

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Dawn of the gene-editing age

PAGE 155



EVERYWHERE

CONSERVATION

A WORLD OF TWO HALVES

E. O. Wilson's vision for an
Earth shared with nature

PAGE 170

PLANT BIOLOGY

FLOWER ARRANGEMENT

An attractant/receptor pair
driving pollen-tube growth

PAGES 178, 241 & 245

GROUP DYNAMICS

THE RIGHT SIZE FOR A LAB

The skills mix and head
count needed for success

PAGE 263

NATURE.COM/NATURE

10 March 2016 \$10

Vol 531, No. 7593

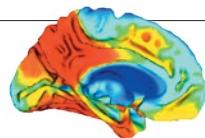


THIS WEEK

EDITORIALS

WORLD VIEW Help the people who helped discovery happen p.141

ELECTRONICS Researchers make stretchy skin that glows p.142



INSIDE VIEW Imaging technique shows ageing protein p.143

Food processing

A recreation of how early humans managed to eat a diet of meat hundreds of thousands of years before they had fire to cook it with, shows an ingenious use of tools to cut down on chewing time.

You are what you eat. Not only that, but you are what your ancestors ate, when they ate it, and what they did to it first. One of the many peculiarities that set humans apart from other animals is that eating is more than just stuffing something into our mouths.

True, the human diet is astonishingly eclectic, but this wide range is tempered by elaborate preparation. No other animal, for example, exposes prospective food items to prolonged heating, a habit we call 'cooking'. It's now generally thought that cooking was central to the evolution of modern humans, prompting a massive reduction in tooth size and chewing muscles, alongside a marked increase in available nutrients, more time to spend doing other things besides chewing, and even an expansion of the brain.

There is — as always — a catch. Cooking requires fire, and there is scant evidence for the regular use of fire before around 500,000 years ago. *Homo erectus*, the first hominin to even begin to approach modern humans in stature, brain size and masticatory apparatus, appeared around 1.5 million years earlier than that. *Homo erectus* was a regular carnivore, a habit that has stayed with us and is believed to be necessary to our modern diet (see *Nature* 531, S12–S13; 2016).

How did *H. erectus* manage to consume meat without cooking it? As Katherine Zink and Daniel Lieberman explore in a paper online in *Nature* (see <http://dx.doi.org/10.1038/nature16990>), raw meat is tough and practically impossible to break down into swallowable pieces just by chewing it. Side orders of roots and tubers can be crunched, but only if you are prepared to put in the hours. A lot of hours. About 40,000 chews a day, which, at a ruminative rate of 1 chew per second, adds up to 11 hours. That's almost a whole day gone, just chewing. That's no issue for many baseball players or football managers, perhaps, but *H. erectus* had better things to do.

The new study squares the circle by showing that tools equivalent to knives, mortars and pestles entered the kitchen a long time before the oven. Stone tools date back to at least 3.3 million years ago (S. Harmand *et al.* *Nature* 521, 310–315; 2015). A freshly struck flake of stone makes short work of slicing raw meat into morsels, and a lump of rock can be used to pound roots and tubers into a paste.

Work with people today has put numbers on these gains. When meat is sliced and roots are pounded, a prehistoric diet of 2,000 kilocalories per day (one-third raw goat and two-thirds raw yams, carrots and beets) can be achieved with 2.5 million fewer chews a year than if the items are unprocessed. That's an entire month spent not chewing — presumably enough to explain the reduction in tooth size and masticatory muscle mass of *H. erectus* compared with earlier, more masticatory species, as well as the increase in brain size allowed by the release of more nutrients. And what does one do with one's mouth when not chewing? One talks a lot, of course. Preferably to other people.

Our ancestors probably also ate fruits and berries, fish and shellfish, nuts, bone marrow, liver and brains, all of which are highly nutritious. But some of those foods need a deal of slicing and pounding to get at.

Nuts have hard shells, as do shellfish, by definition; marrow and brains require (there is no delicate way to put this) the smashing of bones and skulls. Many animals are known to use simple tools to acquire food of that sort. But the release of nutrients from muscle by an animal with teeth more suitable for crushing than slicing required the application of some early food technology.

"Raw meat is tough and practically impossible to break down into swallowable pieces just by chewing it."

Cooking, when it came, enabled yet more efficient nutrient release, and provided other benefits such as the killing of any harmful parasites that raw meat might contain, as well as the gathering of sociable people round a hearth to swap gossip, watch celebrity chefs on TV and share pictures of their cats on the Internet, if only as a way of using up all that time not spent chewing the fat. But cooking did not start this. It merely accelerated a culinary tradition already millions of years old. ■

Who ordered that?

An unexpected data signal that could change everything has particle physicists salivating.

Physicists at the Large Hadron Collider (LHC), the giant particle-physics experiment near Geneva, Switzerland, have searched for many possible subatomic particles and novel phenomena. They have tried to recreate dark matter, reveal extra dimensions of and collapse matter into microscopic black holes.

But the possibility of an electrically neutral particle that is four times heavier than the top quark — the current heaviest — and that could decay into pairs of photons has apparently never crossed anybody's mind. No theorist has ever predicted that such a particle should exist. No experiment has ever been designed to look for one.

So when, on 15 December last year, two separate teams at the LHC independently reported hints of such a particle (see *Nature* <http://doi.org/bc4t>; 2015), the reaction of many experts was similar to that of US physicist Isidor Isaac Rabi when the muon, a heavier relative of the electron, was discovered in 1936: "Who ordered that?"

If the particle exists, the implications would be enormous. Precisely because it is so unexpected, it could be the most important discovery in particle physics since quarks — the elementary constituents of protons and neutrons — were confirmed to exist in the 1970s. Perhaps it would be the biggest deal since the muon itself.

The evidence so far is scant, however. It amounts to a few too many

pairs of γ -ray photons produced with combined energies of 750 giga-electronvolts when the LHC smashes protons together. The fact that two separate detectors spotted it at almost exactly the same energies gives some hope, but anomalous signals such as this often show up in experiments only to later vanish back into the noisy background.

Still, people at CERN, the European particle-physics lab that hosts the LHC, have scarcely talked about anything else since. And theoretical physicists around the world have gone into overdrive: more than 200 papers have been posted online with theories that could explain the particle. One possibility is that it could be a heavier cousin of the Higgs boson; another, even more tantalizing one, is that it is a type of graviton, the particle hypothesized to carry the force of gravity. If so, it could point to the existence of extra dimensions of space beyond the familiar three.

Some have discounted the outburst of preprint articles as merely an attempt by authors to rake up citations. One physicist has even done a quantitative comparison of this spike in activity with other fads that have come and gone in the past (see M. Backović Preprint at <http://arxiv.org/abs/1603.01204>; 2016), charting theorists' initially exploding, then fading, interest. But describing theorists' interest as 'ambulance chasing' is a bit unfair. To paraphrase Albert Einstein, if people knew what they should be looking for, it wouldn't be called research.

And particle physicists' excitement is understandable, if tempered by caution. For decades, their field has been finding evidence for the standard model of particle physics, a collection of theories that was put together in the 1970s and has been more successful than anyone expected. The current generation of young physicists was not even born when particle accelerators produced their last genuinely surprising results. Meanwhile, searches for physics beyond the standard model have so far come up empty — at accelerators such as the

LHC but also in many tabletop experiments and at detectors built underground or sent into space to look for dark matter. The most notable exception to the standard model's standard fare has been the discovery, beginning in 1998, that the elementary particles called neutrinos spontaneously oscillate between their three known types, or flavours — something that the original version of the standard model had not predicted. That breakthrough earned two physicists a well-deserved Nobel Prize last year.

"The LHC is now providing the opportunity of a lifetime to break entirely new ground."

The LHC is now providing the opportunity of a lifetime to break entirely new ground. In 2015, it restarted after a long shutdown that brought the energies of its collisions to a record 13 teraelectronvolts, from 8 TeV. This has put much more massive particles in reach — if any exist — but it will be the last substantial jump in collider energies in a generation. More-powerful machines, if they ever see the light of the day, will take decades to plan, develop and build.

The good news is that whether the new particle exists or the data bump is a statistical anomaly is not a question that will leave us hanging for long. The LHC experiments had time to observe only relatively few collisions in their first 13 TeV run last year, before the experiment shut down for its winter recess.

At a meeting in the Italian Alps that starts on 12 March, LHC researchers might present fresh analyses of those data that could provide more clues. And the machine will begin to collect vastly more data in April. If the bump seen last year was an anomaly, it should go away by the summer. If not, stay tuned for some interesting announcements at the next round of conferences. ■

Gene intelligence

The risks and rewards of genome editing resonate beyond the clinic.

Last month, one of the top intelligence officials in the United States warned that genome-editing technology is now a potential weapon of mass destruction. Techniques such as the emerging CRISPR–Cas9 system, US director of national intelligence James Clapper warned in an annual threat-assessment report to the US Senate, should be listed as dangers alongside nuclear tests in North Korea or clandestine chemical weapons in Syria (see go.nature.com/jxuyev).

The headline message might scream 'overreaction' — and indeed most serious science commentators seem to have assumed as much and ignored Clapper's hyperbole — but the terms he used to describe the technology seem uncontroversial. The US spooks describe the "broad distribution, low cost, and accelerated pace of development" of gene editing, and say that its "deliberate or unintentional" use could have "far-reaching economic and national security implications".

"Research in genome editing," the threat assessment continues, "increases the risk of the creation of potentially harmful biological agents or products." And Clapper, naturally, points the finger at science in nations "with different regulatory or ethical standards than those of Western countries". But for a glimpse of just how far-reaching the "deliberate or unintentional" use of gene editing could be, he need only look over his shoulder.

Last year, scientists in California reported that they had used gene editing (together with another new biotechnology called gene drive) to introduce a mutation

that disabled both normal copies of a pigmentation gene on a fruit-fly chromosome. The change made the insects turn pale yellow — as did their offspring, their offspring's offspring and so on. The change was so powerful that, had any of the California flies escaped, it has been estimated that somewhere between one in five and one in two of all the fruit flies in the world would be yellow today. The flies did not escape — but then, weapons of mass destruction are a political problem because they exist, not because they are deployed.

Clapper was anxious about the implications of gene editing because of its dual-use possibilities. But a binary outcome is inadequate for describing the spectrum of ways in which the CRISPR–Cas9 system is changing science and could benefit scientists and the public. In a special issue this week, we examine some of these (see page 155).

Much of the early attention has focused on the prospect of human-embryo modification. The issues that such 'germline' changes could raise for current and future generations have, rightly, been intensely debated. But the uses of CRISPR–Cas9 with early promise are those in laboratories, not clinics — and in human somatic (non-reproductive) cells, bacteria, viruses, animals and plants, not in human germ cells. A pair of News Features starting on page 156 explores these scenarios.

Genome editing is a science for which the alarm about how it could go wrong has largely lagged behind the hype over what good it could achieve — at least before Clapper had his say. And much of the hype has come from those in the know. The speed at which the biological community has adopted gene editing, and the range of applications that it is being used for, speak volumes about its potential. The possibilities — human–animal chimaeras for organ transplants, climate-change-proof crops, eradication of disease vectors — seem endless.

Among the many unknowns that swirl around the future of gene editing is the reaction of the wider public. To their credit, some scientists and organizations are making attempts to foster openness and discussion, on the topic of gene drives, for instance. It is crucial that these deliberations continue, and that such environmental issues are kept scientifically and ethically distinct from concerns relating to clinical applications. ■



CRISPR EVERYWHERE

A *Nature* special issue
nature.com/crispr



Support communities involved in disease studies

Lack of continued help for poor families involved in Huntington's-disease research has sown resentment and mistrust, says Ignacio Muñoz-Sanjuan.

After decades of research, a genetic therapy for Huntington's disease is being tested in clinical trials. Sponsored by Swiss pharmaceutical firm Roche and US-based Ionis Pharmaceuticals, this trial targets the gene that causes the disease. If the new treatment works, it could offer a way to halt progression of this genetic disease — an awful neurodegenerative disorder that attacks mainly the brain. Huntington's is caused by a single gene mutation transmitted in a dominant fashion, so a child has a 50% chance of inheriting the condition if one of their parents carries a single copy of the defective gene.

Before the human genome sequence was available, the study of Huntington's disease relied on the help of generations of people in Venezuela, where it is highly prevalent. They donated samples of skin, blood and semen, and handed over organs of their deceased relatives, including their own children. Yet what have these communities got in return? Despite efforts by pioneering scientist and Huntington's advocate Nancy Wexler, who led the research in the Maracaibo region for over two decades and founded a clinic there, they have received little or no benefit from the research they enabled. Because of inaction by local governments, they largely lack access to genetic diagnosis and counselling, and have inadequate medical care and scant legal protection.

Does the biomedical community have a moral responsibility to ensure sustained support for people who were crucial to its research? I argue that it does. As a scientist dedicated to treating Huntington's disease, I struggle with the knowledge that the current quality of life of those affected is deplorable. I have seen people shunned and neglected by their relatives, sitting alone in darkened rooms, devoid of medical or social support. I have met the children of those affected, who are afraid of what will become of them. Tragically, suicide is common.

Some of the largest clusters of Huntington's disease in the world nestle in Maracaibo townships, especially Barranquitas and San Luis, where roughly one third of families have a history of the disease. Wander the streets of these shanty towns, and you will find symptomatic patients on every street corner; to the uninitiated, their numbers are staggering. Many other families with Huntington's live in similar conditions elsewhere in Latin America, particularly in Colombia, Brazil and Peru.

Many of the people I met there now resent and distrust scientists. They had hoped for treatments, and had expected help with palliative medication and improved living conditions. At the very least, they wanted feedback on how their contribution had helped.

Research, and especially basic research, is fundamentally disconnected from the realities of vulnerable populations. Is it unreasonable to expect investigators and their institutions to assume some responsibility for ensuring

adequate care for volunteers and their quality of life? Perhaps studies in vulnerable populations should not be conducted at all, unless a comprehensive, long-term plan is drafted in cooperation with the research institutions involved and local and national governments.

At a minimum, and as described in 2002 by the Council for International Organizations of Medical Sciences and the World Health Organization, sponsors have a responsibility to ensure that people recruited for research from vulnerable populations "will ordinarily be assured reasonable access to any diagnostic, preventive or therapeutic products that will become available as a consequence". In the case of the ongoing Colombian trial for familial Alzheimer's disease, sponsored by Roche and California company Genentech, patients participating in the study have been guaranteed access to the medication. But this is not enough. Sponsors of drug trials should also support development in the wider community.

What most infuriates the people in the Latin American clusters of Huntington's disease, is that they still lack ready access to the genetic tests that could tell them whether they or their children will develop the disease.

Here, all scientists can help. The biomedical community can lobby and pressure national governments to include Huntington's disease in legislation on rare disorders, which guarantees access to tests and treatments, and then enforce these regulations. Although some legislative framework exists in some of the countries involved, it is hard to access, particularly for poor people.

Governments need to offer free genetic tests to everyone at risk, and to provide adequate genetic and psychological counselling, and recognition of their disease status, even in remote communities. To do this efficiently, a proper census of communities with suspected cases of Huntington's disease is necessary, because many of these communities are unknown to government institutions.

Without support, the cases of Huntington's disease in these communities will increase and create an even worse public-health issue. Governments can develop effective family-planning and gene-carrier identification programmes to curtail the prevalence of the disease. Such an approach has been successful in, for example, diminishing the incidence of β-thalassaemia on the Italian island of Sardinia. And because Huntington's disease affects only a few thousand people in each country, there is an opportunity to make a real difference.

The clinical trial of the new therapy is terrific news. But we must not forget or ignore the needs of those who made it possible. ■

NATURE.COM

Discuss this article
online at:
go.nature.com/s7bzj2

Ignacio Muñoz-Sanjuan is vice-president of biology at the CHDI Foundation and the founder of Factor H, a project to help Latin American communities affected by Huntington's disease.
e-mail: ignacio.munoz@chdifoundation.org

RESEARCH HIGHLIGHTS

Selections from the scientific literature

CLIMATE

Worst drought in centuries

The 15-year drought that ended in 2012 in parts of the Middle East was probably the worst dry spell in the region for 900 years.

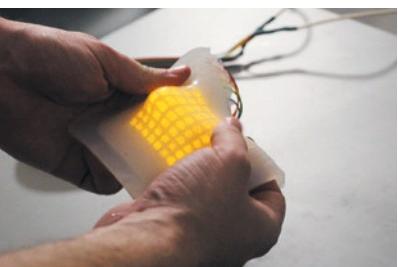
Benjamin Cook at the NASA Goddard Institute for Space Studies in New York and his colleagues analysed tree-ring patterns from 1100 to 2012 to estimate drought variability in the Mediterranean. Summer droughts of similar magnitude to those that have hit the western Mediterranean and Greece in recent decades did previously occur. But the researchers found an 89% likelihood that the 1998–2012 drought in the part of the eastern Mediterranean called the Levant was the driest since 1100.

Climate change will probably increase the risk of drought in the region, potentially aggravating sociopolitical and economic disruption in crisis regions such as Syria, the authors say. *J. Geophys. Res.-Atmos.* <http://doi.org/bcz2> (2016)

ELECTRONICS

Stretchy artificial skin that glows

Inspired by the octopus, researchers have developed an artificial skin that responds to pressure and emits light when stretched.



JUNGBAE PARK

BEHAVIOURAL ECOLOGY

Fungus makes tree frogs sing

A fungal disease that is devastating many amphibian populations around the world causes some infected tree frogs to sing more, even though they don't show other symptoms.

Amphibians are threatened by a global pandemic of chytridiomycosis, which is caused by the chytrid fungus *Batrachochytrium dendrobatidis*. Deuknam An and Bruce Waldman from Seoul National University recorded the mating calls of male Japanese tree

frogs (*Hyla japonica*; pictured), before testing them for the fungus. They found that infected males tended to call more rapidly, and produce longer calls, than non-infected frogs.

This could be a sign that the fungus is manipulating the frogs' behaviour — longer calls attract more frogs, potentially spreading the disease. Alternatively, the frogs could be mating earlier because of a shortened lifespan.

Biol. Lett. 12, 20160018 (2016)

Rob Shepherd at Cornell University in Ithaca, New York, and his colleagues made the skin (pictured) by combining layers of transparent electrode-containing hydrogels with stretchy silicone sheets embedded with various zinc sulfides. They added light-emitting metal compounds to the zinc sulfides, causing them to emit different colours in response to electrical excitement. The team rolled, folded and stretched the material by nearly 500% without disrupting light emission. And the more the material was stretched, the brighter the light.

The authors incorporated

panels of their material into a crawling soft robot, allowing it to luminesce as the robot undulated and the skin stretched. Pressing on the material altered its capacitance — its stored electric charge — so the researchers say that the skin could have applications in touch-sensitive robotics.

Science 351, 1071–1074 (2016)

GENOMICS

Disabling a gene may not be harmful

People who have non-functioning genes may not always have health problems.

David van Heel of Queen Mary University of London, Richard Durbin of the Wellcome Trust Sanger Institute in Hinxton, UK, and their colleagues sequenced the part of the genome that encodes proteins from more than 3,000 healthy adults whose parents were closely related (often first cousins). The team found that 821 individuals carried rare genetic variants that would be expected to cause the loss of function of certain genes. When the researchers examined the participants' health records, they found no links between the loss-of-function genes and clinical

ROB KURCOBA/CORNELL

effects. One pregnant mother lacked a functional *PRDM9* gene, which is required for fertility in mice, but the non-functioning gene had no impact on her health.

Non-functioning genes in adults may not be as clinically important as previously thought, the authors say. *Science* <http://doi.org/bc3x> (2016)

CLIMATE CHANGE

Climate shift for African farming

Many farmers in Africa may have to change the crops they are growing by the end of this century because of climate change, but for most plants only small areas will be impacted.

Julian Ramirez-Villegas at the University of Leeds, UK, and his colleagues modelled the suitability of sub-Saharan Africa for growing 9 major crops under climate scenarios that would see relatively large increases, exceeding 2 °C, in global temperatures by 2100. For maize (corn) and banana, around 30% of the region will become unsuitable, and for beans, 60% of the land will be unavailable. But for the other six crops — including cassava and yam — the affected area is limited to small pockets that total less than 15%.

The authors suggest that some farmers will initially adapt to climate change through improvements to farming techniques, but will then need to transition to substitute crops or relocate. *Nature Clim. Change* <http://dx.doi.org/10.1038/nclimate2947> (2016)

VIROLOGY

Zika virus infects brain cells

Laboratory-grown human cells that are similar to those in the brains of developing fetuses are rapidly infected and killed by Zika virus.

With the disease now spreading across Latin

America and the Caribbean, researchers are racing to understand Zika virus and its potential link to microcephaly in fetuses. Hongjun Song and Guo-li Ming at Johns Hopkins University in Baltimore, Maryland, along with Hengli Tang at Florida State University in Tallahassee and their team, caused reprogrammed human stem cells to develop into neural progenitor cells, then infected them with Zika virus, which replicated rapidly. After three days, the virus had killed one-third of the cells. Immature neurons were also susceptible to Zika, but to a lesser extent.

Neural progenitor cells could be used to study the virus in the lab and identify treatments, the researchers say.

Cell Stem Cell <http://doi.org/bc3w> (2016)

GENETICS

Genetic link for a monobrow

Researchers have identified ten genetic variants linked to hair traits, including the rate at which hair goes grey and whether a person will have a 'monobrow'.

Previous studies looking at European and East Asian populations have identified genes associated with male-pattern baldness, hair colour and curliness. Kaustubh Adhikari at University College London and his colleagues studied the genomes of more than 6,000 people living in Brazil, Colombia, Chile, Mexico and Peru, categorizing volunteers according to the colour, shape and pattern of hair on their scalp and faces.

They found, for example, that the variant associated with the rate of hair greying is in a gene called *IRF4*, which regulates the production and storage of melanin — the pigment that determines hair, skin and eye colour. A variant of *FOXL2* is linked to eyebrow thickness,

SOCIAL SELECTION

Popular topics on social media

'Creator' paper sparks concern

A paper that attributed the architecture of the human hand to "the proper design by the Creator" has triggered a debate over the quality of editing and peer review at high-volume journals.

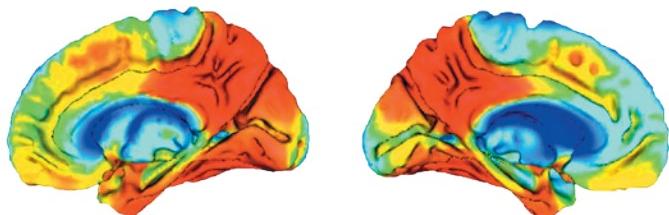
The paper by Cai-Hua Xiong at the Huazhong University of Science and Technology in Wuhan, China, and his co-authors appeared in *PLoS ONE* on 5 January. But it came to prominence last week after its apparently creationist slant was flagged on Twitter, spawning the hashtags #Creatorgate and #HandofGod. James McInerney, who works on computational molecular evolution at the University of Manchester, UK, started the ball rolling with a tweet, saying the paper reveals *PLoS ONE* to be an "absolute joke of a journal". When contacted by *Nature*, Xiong said he was sorry, adding, "We are not native speakers of English, and entirely lost the connotations of some words such as 'Creator'." The journal later posted an online statement saying that it had decided to retract the paper. "Our internal review and the advice we have

received have confirmed the concerns about the article and revealed that the peer review process did not adequately evaluate several aspects of the work."

NATURE.COM

For more on popular papers: go.nature.com/5641cx

PLoS ONE <http://doi.org/bc4c> (2016)



and a *PAX3* variant is associated with the growth of a monobrow.

Nature Commun. 7, 10815 (2016)

NEURODEGENERATION

Ageing protein imaged in brain

A protein that accumulates in the brain with normal ageing as well as with Alzheimer's disease can be tracked using human brain imaging for the first time.

Scientists could previously map the insoluble form of the protein tau in human brain tissue only after death. To follow changes in tau levels and distribution over time, William Jagust at the University of California, Berkeley, and his colleagues used a previously developed molecule that labels tau for positron emission tomography (PET) imaging

(pictured) in living people. Compared with young people, healthy older people had increased tau in the medial temporal lobe, an area involved in memory. Higher levels of the protein predicted a poorer performance on certain memory tasks. Older adults with suspected Alzheimer's disease had the highest levels of tau. Across all older participants, the spread of tau to other brain areas correlated with higher levels of amyloid- β protein, which is also associated with Alzheimer's disease.

The technique could be used to monitor brain health and test drug candidates, the authors suggest.

Neuron 89, 971–982 (2016)

NATURE.COM

For the latest research published by *Nature* visit: www.nature.com/latestresearch

POLICY

Nuclear research

The US House of Representatives approved a bipartisan bill on 29 February supporting nuclear-energy research and development. The legislation, which seeks to promote engagement between the US energy department and private firms, would bolster existing research activities and direct the agency to create a new open-access nuclear-science facility. The US Senate has approved its own nuclear-research legislation as part of a larger energy bill, although progress has been stalled by disputes over federal funding to help the city of Flint, Michigan, to deal with lead contamination in its water supply.

EVENTS

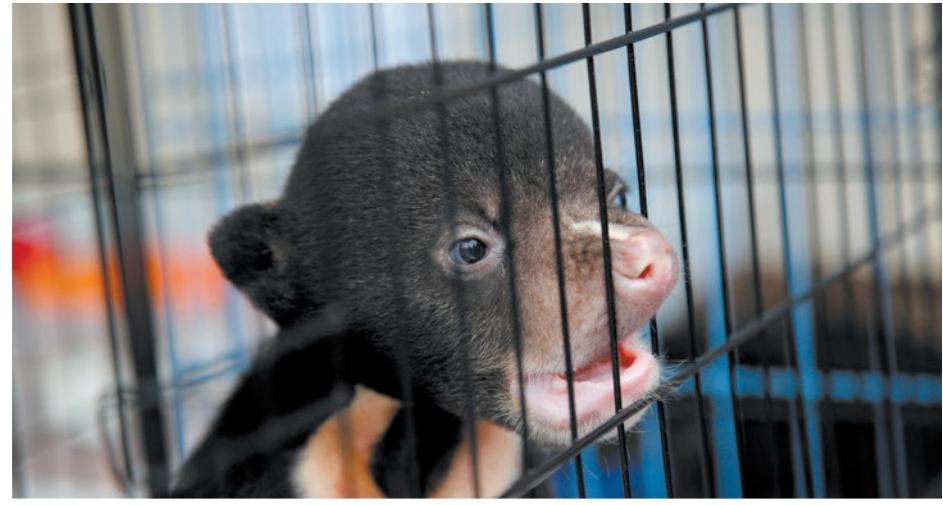
Quake warning fails

Tsunami warning buoys off the coast of Indonesia all failed when a powerful undersea earthquake hit on 2 March, 800 kilometres west of Sumatra. In the wake of the deadly 2004 Indian Ocean tsunami, scientists deployed the array of buoys to observe unusual changes

NUMBER CRUNCH

8,000

The number of Sumatran orangutans living in the wild in addition to the number previously thought to exist there. A survey has found 14,600 of the critically endangered great apes in their natural habitat, more than previous surveys estimated. Some 4,500 animals are still threatened if planned deforestation continues.



DAMIR SAGOLJ/REUTERS

Concern over Facebook wildlife trafficking

Facebook has become a major marketplace for illegal trading in wildlife on mainland Malaysia, according to the non-profit organization TRAFFIC, headquartered in Cambridge, UK. A 3 March report showed that 80 live species were traded using 14 closed Facebook groups in just 50 hours of monitoring over 5 months. Half of the species were on

sale illegally, including sun bears (*Helarctos malayanus*; pictured) and white-handed gibbons (*Hylobates lar*). Illegal wildlife trading had been considered less of a problem in peninsular Malaysia than in most of southeast Asia because of a lack of physical markets. TRAFFIC says that the findings may indicate a global shift in wildlife crime.

in water movement and sea level. But when the latest magnitude-7.8 quake hit, all 22 buoys were down. Officials with Indonesia's Disaster Mitigation Agency blamed vandalism and a lack of funds for maintenance. Authorities issued a precautionary warning on the basis of the quake's location and strength. The warning was lifted when no tsunami was triggered.

Poaching concern

Elephant populations in Africa are still being driven down by poaching, according to the latest data from the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). A 3 March report from CITES notes that poaching continues to decline from its 2011

peak, but says that poachers probably took more animals from wild populations last year than were replaced by births. John Scanlon, the CITES secretary-general, warned in a statement that African elephant populations "continue to face an immediate threat to their survival", especially in Central and West Africa, where poaching is highest.

Bear boom

The US Fish and Wildlife Service has proposed removing the grizzly bear (*Ursus arctos horribilis*) from the list of threatened species under the Endangered Species Act. The change would apply to the bear population in and around Yellowstone National Park in the US northwest, where numbers have increased from

as few as 136 in 1975 to more than 700 today. The proposal, announced on 3 March, would transfer authority for continuing recovery efforts to state management and Native American tribes, although federal biologists would continue to monitor the bears' progress. Environmentalists are expected to challenge the proposal.

ExxonMobil probe

The US Department of Justice has referred congressional requests for a probe into Texas-based oil company ExxonMobil to the Federal Bureau of Investigation (FBI). California representatives Ted Lieu and Mark DeSaulnier requested an investigation in October 2015 into whether the company broke the law by "failing to

disclose truthful information" on climate change to investors and the public. A 12 January response from the Department of Justice, revealed publicly on 2 March, said that the FBI will determine "whether an investigation is warranted".

Climate reporting

US broadcasters reported on climate change less in 2015 than in 2014. A 7 March survey by non-profit organization Media Matters for America tracked coverage by the ABC, CBS, NBC and Fox networks on their evening and Sunday news shows. In 2015, they devoted a total of 146 minutes to climate reports, 5% down on 2014. ABC devoted the least time, at just 13 minutes. PBS, which was considered separately, featured 58 climate-change-related segments in its nightly news show in 2015, much more than the other channels, which aired a total of 48 segments between them.

FACILITIES

Japanese collider

Japan's newest accelerator, SuperKEKB in Tsukuba, has circulated its first particles. The KEK laboratory announced on 2 March that in February, beams of electrons and their antiparticles, positrons, travelled at close to the speed of light around the accelerator's 3-kilometre ring



at separate times. Next year, SuperKEKB will run both beams together and smash the electrons and positrons inside the accelerator's Belle II detector (pictured), to study the asymmetries between matter and antimatter. The experiment is designed to produce collisions at a rate at least 40 times faster than the original KEKB.

PEOPLE

Pachauri charged

Police in New Delhi have filed charges against Rajendra Pachauri, executive vice-chairman of The Energy and Resources Institute (TERI), also in New Delhi, and former chairman of the Intergovernmental Panel on Climate Change. The charges, reportedly made on 1 March, follow a sexual-harassment complaint by a female TERI researcher in February 2015 and include: making physical

contact, using unwelcome and "sexually coloured" remarks, stalking using e-mails and text messages, and criminal intimidation. Pachauri has denied all the allegations against him.

FUNDING

Funding fears

The UK Research Councils have warned that government budget plans unveiled on 4 March "create pressures ... that may have an impact on some existing commitments". The government announced in November (see *Nature* 528, 20; 2015) that it would continue to place a 'ring fence' around the core science budget to protect it from cuts, holding it at £4.7 billion (US\$6.7 million) in real terms. But the budget breakdown of planned allocations to the seven UK councils shows that, although the overall budget is increasing, the councils that

COMING UP

12–15 MARCH

A conference on cancer as an evolving and systemic disease convenes in New York City.

go.nature.com/77upv8

14 MARCH

The European Space Agency and Russia's space agency, Roscosmos, launch the trace gas orbiter to Mars, part of the ExoMars project. The instrument will measure methane and other gases in the planet's atmosphere.

go.nature.com/s7nbe

14–18 MARCH

The World Bank holds its annual land and poverty conference at its headquarters in Washington DC.

go.nature.com/gm3ye3

support biological, physical, environmental, economic and humanities work will have smaller resource budgets by 2019. The agencies for medical research and facilities will see their funds increase.

Illumina chief

The genome-sequencing giant Illumina announced on 7 March that Jay Flatley is stepping down as its chief executive — a post he has held since 1999. Under his tenure, the San Diego biotechnology firm has grown from a 30-person enterprise to a US\$2.2-billion behemoth employing more than 4,800 people. More than 90% of all DNA sequencing data is now produced by Illumina machines. Flatley will be replaced by Illumina president Francis deSouza, and will become executive chairman of the company in July.

NATURE.COM

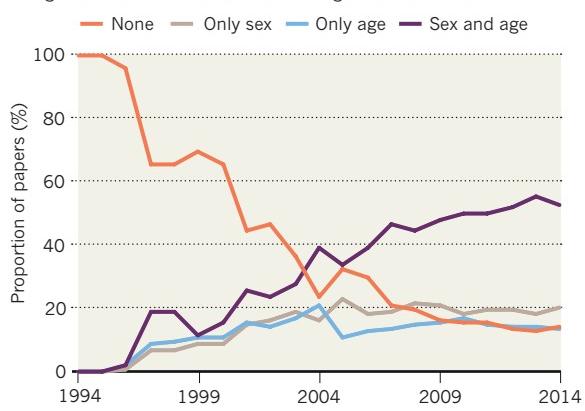
For daily news updates see:
www.nature.com/news

TREND WATCH

In 2014, only around 50% of research papers recorded both the sex and age of the animals used in mouse studies. These details are needed for others to assess and reproduce the work. The analysis used software to trawl through some 15,000 open-access papers published between 1994 and 2014. Recording improved through the 1990s and 2000s, but standards plateaued after 2010 — despite the launch that year of a voluntary reporting checklist called the ARRIVE guidelines. See go.nature.com/vmehx5 for more.

SLOW PROGRESS IN MOUSE STUDIES

A text-mining analysis suggests that, by 2014, 53% of experiments using mice recorded both the sex and age of the animals.



NATURE | NEWS

'Open-hardware' pioneers push for low-cost lab kit

Conference aims to raise awareness of shared resources for building lab equipment.

Elizabeth Gibney

08 March 2016



[Print](#)

Univ. Tübingen & Christoph Jäckle

DIY labware: a 3D printable micropipette (left) and FlyPi, a 3D printable open-source platform for optical microscopy (right).

Few scientists know that, instead of buying their lab equipment, they can often build it much more cheaply — and customize their creations — by following 'open-hardware' instructions that are freely available online.

Fifty enthusiasts who gathered last week at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, are hoping to remedy researchers' lack of awareness about open science hardware. At the first conference dedicated to the field, they met to compare creations — and to thrash out a road map to promote the widespread manufacturing and sharing of labware. "We want open hardware to become a normal part of the scientific process," says Shannon Dosemagen, a co-organizer of the conference who is executive director of the non-profit citizen-science community Public Lab.

Proponents of open hardware — named by analogy to 'open software' in computer science — have already created free online designs for dozens of pieces of labware, taking advantage of manufacturing technologies such as 3D printers and laser-cutting machines. They argue that sharing designs for others to adapt can vastly accelerate the progress of science. But this share-all do-it-yourself (DIY) philosophy is yet to become mainstream. "The majority of scientists are still waiting to get involved," says Joshua Pearce, an engineer at Michigan Technological University in Houghton, who two years ago published a book for scientists on how to create a low-cost lab.

Low-cost lab kit

The open-hardware movement can already point to much success in science, says conference co-organizer Jenny Molloy, who coordinates OpenPlant, a synthetic-biology centre at the University of Cambridge, UK. Citizen-science projects, schools and researchers who lack money to buy expensive equipment have been particularly quick to adopt it. In 2009, for example, Irfan Priyambada, a microbiologist at Gadjah Mada University in Yogyakarta, Indonesia, was able to equip his lab with tissue-culture

Related stories

- Environmental science: Pollution patrol
- Mobile science
- Laboratory equipment: Cut costs with open-source hardware

[More related stories](#)

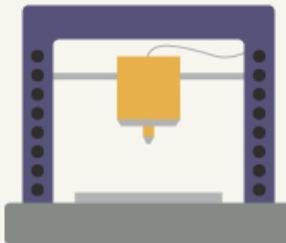
hoods and microscopes for less than 10% of their commercial price, using designs posted by a life-sciences-community platform called Hackteria.

HOW TO MAKE A... DIGITALLY CONTROLLED SYRINGE PUMP

To deliver liquid at precise rates and volumes — whether for mixing reagents, pumping cells into culture or spinning polymers, every lab needs a syringe pump. Follow *Nature's* guide to making your own.

1 Pick a recipe for a syringe pump online (go.nature.com/slzvlh).

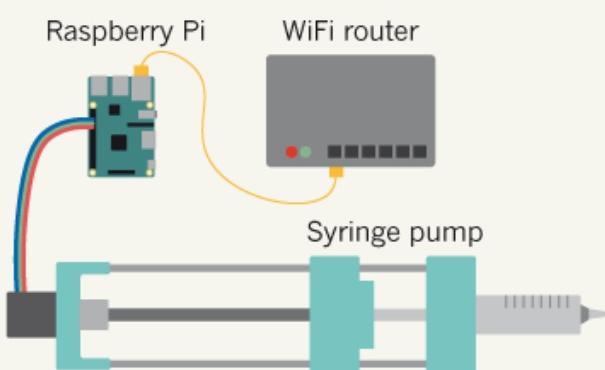
2 Download 3D-printer files and print out plastic components.



3 Buy remaining parts, such as a Raspberry Pi controller, Ethernet cables, motor, bearing, rods and screws.

4 Allen key and drill bit at the ready: the syringe can be assembled in less than an hour.

5 Install software on Raspberry Pi, plug into syringe motor, and connect to computer via router to calibrate.



Open-source price: <US\$100 for single; about \$150 for a dual syringe pump

Proprietary price: \$250–\$1,500 for single; \$1,800–\$2,600 for dual pump

©nature

Online designs have been created for a wide range of labware, from DNA-amplifying PCR machines to fluorescence imaging microscopes (see 'How to make a ... digitally controlled syringe pump'). (Molloy says that the basic principles behind a lot of labware are not patented, so intellectual-property conflicts are rare.) For some kit — such as scanning tunnelling microscopes — the fabrication process is too complex to take place in the lab, but Pearce thinks that these, too, will eventually become open source. And because these blueprints are openly shared — allowing anyone to critique and improve them — the quality of equipment is often at least as good as or even better than what is available commercially, he says.

For researchers, this ability to tinker with equipment is the main advantage of open-source sharing. "If it's open source, I can adapt it and fix it. That's most important to me," says Tobias Wenzel, a biophysics PhD student at the University of Cambridge.

Quality assurance

But other scientists' reluctance to dive into DIY may stem from doubts about whether open hardware can faithfully produce the validated, standardized performance of commercial equipment. Too often, the documentation that accompanies designs — intended to calibrate the equipment's performance against known standards and describe its use — is unclear or incomplete, conference attendees heard. A community-standard or best-practice guide could use a checklist to ensure that designers cover all the necessary bases, says Wenzel. "It needs to be something that says: 'if you follow this procedure, this will work and you'll be able to get high precision, high accuracy and low error,'" says Pearce.

The problem is that sharing work in enough detail for anyone else to follow takes time and effort, but provides little formal scientific credit. "It's one thing to build something for one's own research, but to make it so it's easy for others to replicate is much more difficult," says Ryan Fobel, an engineer at the University of Toronto, Canada, who helped to develop an open-source platform for doing biology and chemistry on a chip, known as DropBot.

To this end, at the Geneva conference researchers debated ways to assign credit to the designers of open hardware. Some would like to see a citation system for designs, or want journals to publish more research papers that outline designs. A central repository for open science hardware might help: CERN hosts a repository for electronics open hardware, and the US National Institutes of Health has a 3D-printing repository with a labware section. But no single repository collates everything.

Because many scientists won't want to build devices themselves, taking open hardware mainstream will need to involve non-profit

organizations and companies that can supply the kit, notes Francois Grey, a physicist at the University of Geneva and conference co-organizer. Firms such as OpenTrons in Brooklyn, New York, which makes automated pipetting systems, already both design open-source lab equipment and sell ready-made kit built from open-source designs. But because such companies give away their designs, figuring out a solid business model is a challenge, adds Javier Serrano, an engineer at CERN who helped to pioneer the lab's Open Hardware Licence, which allows developers to ensure that all future modifications are documented and shared.

Companies might make money by providing support for open hardware, or by conducting quality-assurance checks and validation tests that allow them to offer warranty-like guarantees on products, Pearce suggests. And a collection of success stories might also help scientists to convince their institutions — which may be accustomed to patenting in-house inventions — of the value of forming open-hardware spin-offs, adds Molloy.

Pearce says that he dreams of a day when every published scientific article will instruct its readers not just on experimental methods, but also on how to build the equipment that the study requires. It's something that will need the cooperation of funders to become a reality. Existing large-scale equipment grants tend to pay for single instruments, but Pearce would like to see the money spent on open-source hardware, which he says could bring down prices and — over time — improve designs.

Nature 531, 147–148 (10 March 2016) doi:10.1038/531147a

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

Related stories and links

From nature.com

- **Environmental science: Pollution patrol**
07 January 2015
- **Mobile science**
01 October 2014
- **Laboratory equipment: Cut costs with open-source hardware**
29 January 2014
- **Homegrown labware made with 3D printer**
16 April 2012
- **Blog post: Promoting shared hardware design**

From elsewhere

- **Gathering for Open Science Hardware 2016**
- **Open Labware**
- **Teaching and Research in Natural Sciences for Development in Africa**

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be redirected back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

3 comments

[Subscribe to comments](#)

- 
- Andrej Mosat • 2016-03-09 08:29 PM
- Great overview of open hardware and open lab initiatives. Disclaimer: I make it a living building laboratory spectrophotometers for others. It was customary in Central Europe before 1989 to provide schematics with your purchased TV or radio, so you could repair it at home. I like to see the concept of openness as a means to create and advance knowledge, engineering and reduce waste. This, especially the reducing waste part, is indeed associated with increased costs, not decreased. As long as the device performs similarly to closed-source counterparts, the price should be the same, or slightly higher, because the development effort is the same, if not greater. It makes sense to identify that competitive advantage, such as best performance to price ratio. The unbeatable advantage of open hardware lies indeed in the option

to modify, tinker and advance the design in unforeseen ways. As an example, if I build and sell open spectrometers, I expect the scientists not to waste time looking for ways to make the design cheaper or spend months on replicating the same device, but rather advance it in new applications like Raman spectroscopy, hyperspectral imaging or remote sensing. And this is exactly the trend I observed in the last three years. The technology, tools and knowledge allows almost everyone to start iterating on open hardware. It would be great to have a phylogenetic tree of designs akin to github, but for physical things instead.



Ryan Fobel • 2016-03-09 04:16 PM

Nice article, but I for one would like to see less of a focus on cost. Yes, driving down costs democratizes access to the tools, but there are more compelling reasons for openness. Open hardware enables widespread adoption of new tools that may not otherwise exist because they do not have a sufficient commercial market (especially at early stages of development). It also allows researchers to understand, modify and improve the tools that they depend on for their research. In general, open hardware represents a better alignment between the interests of the scientific community and the people who fund it (i.e., the public). One could argue that we should be willing to pay a premium for openness as it represents added value.



Laura O'Hara • 2016-03-09 01:19 PM

'Some would like to see a citation system for designs, or want journals to publish more research papers that outline designs'... come on then Nature, maybe it's time put your money where your mouth is and launch Nature Open Hardware!

See other News & Comment articles from *Nature*

Nature ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER

NATURE | NEWS

How the US CRISPR patent probe will play out

Decision could determine who profits from the gene-editing technique in future.

Heidi Ledford

07 March 2016



MIGUEL RIOPA/AFP/Getty Images

Emmanuelle Charpentier (left) and Jennifer Doudna (right) seek gene-editing patents.

There is no shortage of optimism about the scientific potential of CRISPR–Cas9, a technique that can precisely alter the genomes of everything from wheat to elephants. But there is a great deal of confusion over who will benefit financially from its use.

On 10 March, the US Patent and Trademark Office (USPTO) will begin an investigation into who deserves the patent on using CRISPR–Cas9 to edit genes. This ‘patent interference’ could determine who profits from CRISPR in coming years.

Already, companies have sprung up to take advantage of the technique in agriculture, industrial biotechnology and the treatment of human diseases. One firm, Editas Medicine in Cambridge, Massachusetts, raised US\$94 million when it went public on 2 February, even though it does not expect to enter clinical trials until 2017.

Nature takes a look at what the interference proceeding entails and what it could mean for the fate of CRISPR–Cas9.

Who's who in the patent interference?

One patent claim comes from a team led by molecular biologist Jennifer Doudna at the University of California, Berkeley, and microbiologist Emmanuelle Charpentier, now at Umeå University in Sweden and the Max Planck Institute for Infection Biology in Berlin. They published a 2012 paper demonstrating that the Cas9 enzyme can be directed to cut specific sites in isolated DNA (M. Jinek *et al.* *Science* **337**, 816–821; 2012), and initiated their patent application on 25 May 2012.



**Nature special:
CRISPR**

Another team, led by Feng Zhang at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, published a 2013 paper demonstrating the application of CRISPR–Cas9 in mammalian cells (L. Cong *et al.* *Science* **339**, 819–823; 2013). Zhang's team began a patent application on 12 December 2012.

Although the Berkeley team filed first, the Broad team submitted its application to an expedited review programme, and was awarded the patent in April 2014. The Berkeley team then requested a patent interference against the initial Broad patent plus 11 related Broad patents. On 11 January, the USPTO granted Berkeley's request.

What is a patent interference?

A relic from the past. Until a few years ago, the United States awarded patents to those who could show that they were the first to invent, rather than simply the first to file the patent. Under that system, when competing inventors claimed to have created the same invention first, the USPTO declared an interference proceeding to determine which deserved the patent.

The United States switched to a first-to-file system in March 2013. But several key CRISPR–Cas9 patents were filed before the change.

What will happen during the patent-interference hearing?

A panel of three USPTO patent judges will hear evidence from both sides to establish which team invented the application of CRISPR–Cas9 for gene editing. Much of the action will be carried out over the telephone or through written documents. But there will probably be some oral arguments, and these could include testimony from the academic inventors.

Patent interferences can be highly technical, says John Conley, a legal scholar at the University of North Carolina in Chapel Hill. “It’s hard for me to cite anything more convoluted in the law than this,” he says. “It’s mind-boggling.” The USPTO panel will probably try to determine not only which team was the first to use CRISPR–Cas9 for gene editing,

Related stories

- Bitter fight over CRISPR patent heats up
- CRISPR, the disruptor

but which conceived of the invention first.

The process could be messy. During the era of 'first-to-invent' patents, some companies kept 'inventor's notebooks': when someone at the firm thought of a new invention, they were to write it down in the notebook and have the entry notarized in case it came into play during future patent disputes. Few academic labs go to such lengths.

- Chinese scientists genetically modify human embryos

More related stories

When will we find out who has won?

The law that did away with the United States' first-to-file policy also introduced changes intended to expedite interferences. But a verdict on the CRISPR patents could still be months, or even years, away. And given the high financial stakes, many expect the losing party to appeal against the USPTO interference decision, further dragging out the process.

Will this be the only CRISPR patent interference?

Not necessarily. In its filings to the Securities and Exchange Commission, Editas Medicine highlighted a potential interference claim by a Seoul company called ToolGen. Having multiple interferences over the same patent is rare, says Conley, but possible.

Is the patent landscape any clearer in Europe?

No. The Broad and MIT team also fast-tracked several of its applications at the European Patent Office (EPO), and has been awarded several patents so far. Doudna's single application is pending.

Although the EPO does not have an interference process, outside parties can formally object to a patent. By 11 November 2015, the deadline for objections to the Broad's first European CRISPR–Cas9 patent, nine parties had come forward — launching an opposition procedure that can take years to resolve.

Once that process is finished, participants can appeal. This adds another four or five years to the clock, says Michael Roberts, a partner at the intellectual-property law firm Reddie & Grose in Cambridge, UK. For this reason, Roberts believes that it will be several years before there is clarity on the earliest CRISPR–Cas9 patents in Europe.

Nature **531**, 149 (10 March 2016) doi:10.1038/531149a

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

Related stories and links

From nature.com

- Bitter fight over CRISPR patent heats up

12 January 2016

- **CRISPR, the disruptor**

03 June 2015

- **Chinese scientists genetically modify human embryos**

22 April 2015

- **CRISPR technology leaps from lab to industry**

03 December 2013

- **Nature special: CRISPR**

From elsewhere

- **US Patent and Trademark Office**

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

1 comment

[Subscribe to comments](#)



S Ananthanarayanan • 2016-03-08 06:21 PM

The story is unclear on some issues. When patents are applied for, there is a notice inviting objections and also a thorough search over international databases to check for prior claims. How did these fail? These are the aspects that would make the story interesting. It is not good enough to just say, "....the Broad team was awarded the patent in April....." There has been a failure here. Patent claims are messy, but they do not run into trouble that lasts years without a reason. (The writer is a journalist -- see <http://www.simplescience.co.in> , piece number 290-- and also a patent attorney)

See other News & Comment articles from *Nature*

Nature ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER

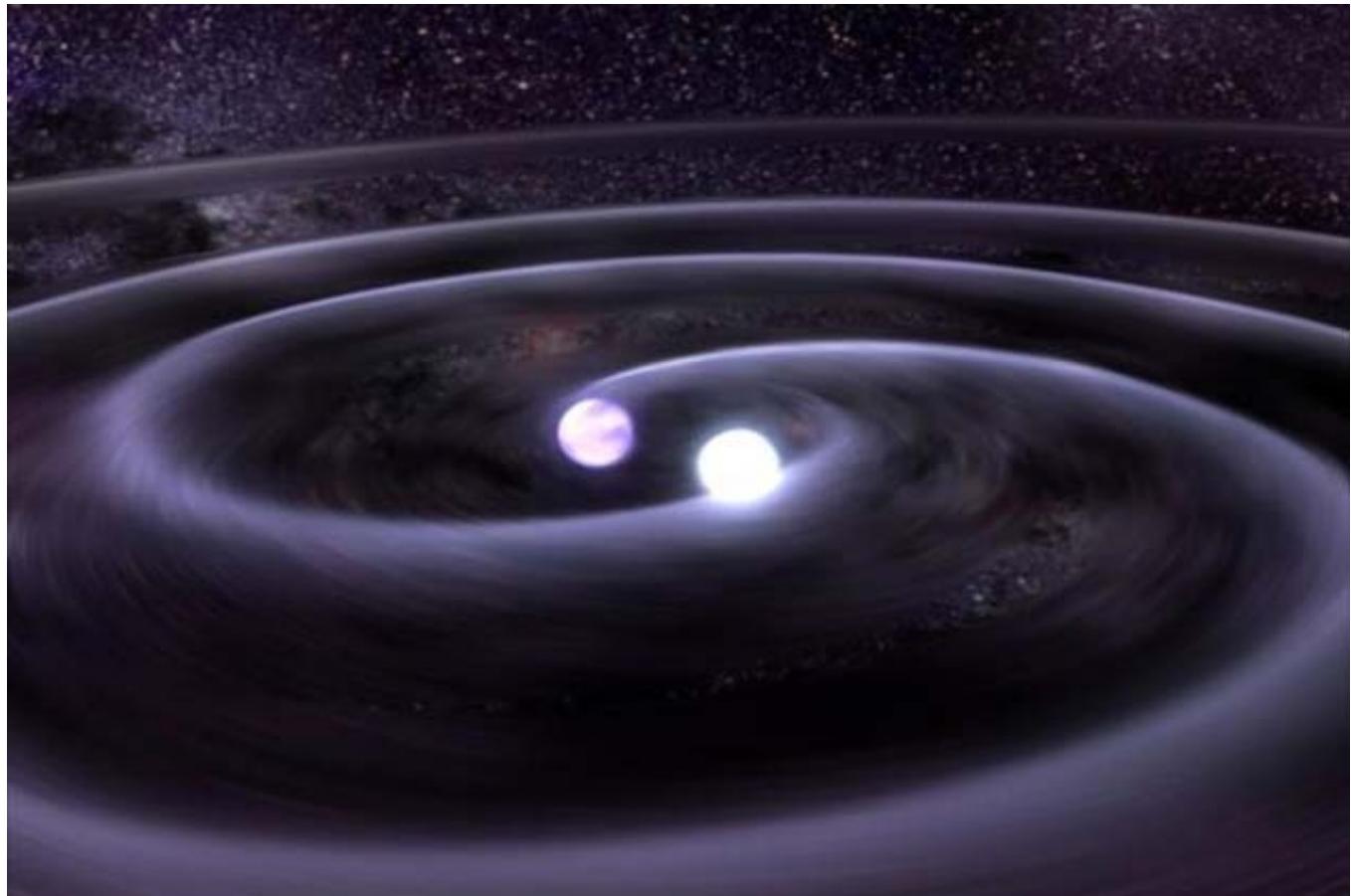
NATURE | NEWS

Chinese gravitational-wave hunt hits crunch time

The pressure is on to choose between several proposals for space-based detectors.

David Cyranoski

09 March 2016



GSFC/D.Berry/NASA

Gravitational waves from the binary star system HM Cancri (artist's impression) would be the target of TianQin, one of China's proposed space-based detectors.

In the wake of last month's historic detection of gravitational waves by a US-led collaboration, a range of Chinese proposals to take studies of these ripples in space-time to the next level are attracting fresh attention.

The suggestions, from two separate teams, are for space-based observatories that would pick up a wider range of gravitational radiation than ground-based observatories can. The most ambitious plan could give China an edge over the leading European proposal to detect gravitational waves from space, but whether a single country can achieve that on its own is unclear. Also under consideration are a possible collaboration between Chinese researchers and the

Related stories

- Successful test drive for space-based gravitational-wave detector
- Einstein's gravitational

European effort, and a cheaper Chinese plan.

Although an Earth-based detector — the US Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) — was the first to confirm a prediction made by Albert Einstein a century ago, launching the field of gravitational-wave astronomy, such detectors can pick up only limited frequencies. Advanced LIGO compares laser light beamed along two perpendicular detector arms to reveal whether one beam has been compressed or stretched by gravitational waves.

Each LIGO arm measures 4 kilometres, but picking up the frequencies that are richest in gravitational waves requires distances of hundreds of thousands of kilometres or more. This can be achieved only in space, where spacecraft equipped with lasers can be positioned at these distances. Space-based detectors also avoid fluctuations in Earth's gravitational field, which can obscure signals.

With such considerations in mind, the European Space Agency (ESA) is pursuing a space-based gravitational-wave detector. One of the Chinese proposals, Taiji, meaning 'supreme ultimate', is to create a more ambitious version of the leading proposal for the European project, which is called eLISA (Evolved Laser Interferometer Space Antenna).

Like eLISA, Taiji would consist of a triangle of three spacecraft in orbit around the Sun, which bounce lasers between each other (see 'China's choices'). The distance between eLISA's components is still under discussion, but current plans suggest it could be 2 million kilometres, says eLISA member Karsten Danzmann of the Max Planck Institute for Gravitational Physics in Hanover, Germany. Taiji's spacecraft would be separated by 3 million kilometres, giving the detector access to different frequencies. Taiji would launch in 2033, slipping in a year ahead of eLISA's current schedule. "If Taiji produces a Chinese version of eLISA, then it will bring China to the frontier," says Yanbei Chen, a gravitational-wave physicist at the California Institute of Technology in Pasadena, who works on LIGO.

waves found at last

- Freefall space cubes are test for gravitational wave spotter

More related stories

CHINA'S CHOICES

Chinese researchers have proposed several ways to detect gravitational waves in space.

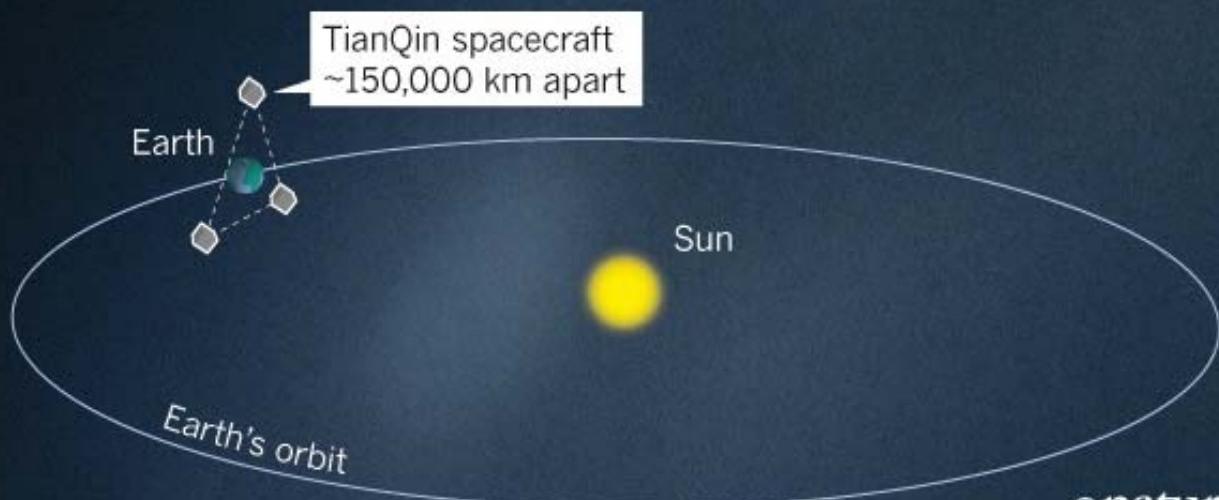
TAIJI

The most ambitious proposal uses three spacecraft in a triangle that orbits the Sun and detects gravitational waves from a range of objects, like Europe's eLISA proposal. The spacecraft are farther apart than in eLISA, giving Taiji access to different frequencies.



TIANQIN

A cheaper proposal puts three craft in orbit around Earth, and much closer to each other than in Taiji. This would target the gravitational waves emitted by HM Cancri, a pair of white dwarf stars.



©nature

SOURCES: eLISA/Wu Yue-Liang/Luo Jun

Gerhard Heinzel, an eLISA physicist also at the Max Planck Institute in Hanover, cautions against a single country going it alone on such a large project. It "is definitely too big — mainly in terms of cost but also resources in terms of scientists and experts in the presence of competing science projects", he

says.

Taiji project leader Wu Yue-Liang, a particle physicist at the Chinese Academy of Sciences' Institute of Theoretical Physics in Beijing, estimates that the project will cost 14 billion yuan (US\$2 billion), roughly twice as much as ESA is budgeting for its gravitational-wave detector.

Second string

A second Chinese proposal, led by Luo Jun, a physicist at the Sun Yat-Sen University campus in Zhuhai, would lower the bar in terms of cost and resources. Called TianQin, a name that refers to the metaphor of nature playing a stringed instrument (a zither) in space, the project has three satellites that orbit Earth at a distance of about 150,000 kilometres from each other. It would cost 2 billion yuan, says Luo.

TianQin would be more limited than Taiji in terms of what it could detect: rather than acting as an observatory for the waves emitted by myriad objects including black holes and neutron stars, it would mainly target a particular pair of orbiting white dwarf stars, called HM Cancri. TianQin's simplicity makes it cheaper and more certain of success, Luo says. The spacecraft could launch in 15–20 years, he adds, around the same time as the Taiji group says that it could launch. Luo thinks that a simpler project is more realistic now, but says that TianQin could lay the groundwork for a Taiji-like project in the future.

Wu Ji, director-general of the Chinese Academy of Sciences' National Space Science Center, says that the TianQin and Taiji teams should merge. "If China decides to have a space gravitational mission, there should be an integrated one, with a new name probably. There is no way to support two missions at the same time."

Both Wu Yue-Liang and Luo are confident that their proposals will move forward to the concrete design phase in the next five years. Taiji currently receives money from the Chinese Academy of Sciences and TianQin from the city of Zhuhai — but both need much more cash. The LIGO discovery could increase their chances of success. The "government will know more the importance of fundamental research" in gravitational waves, says Wu Ji. "China should catch up in this area," he adds.

On 5 March, the Chinese central government released a draft list of 100 strategic projects that will be emphasized in the country's next five-year plan, which includes "a new generation of heavy launch vehicles, satellites, space platforms and new payload" and a "deep-space station".

Chinese researchers could also end up collaborating with Europe. As well as its main project, the Taiji group has outlined the possibility of a direct collaboration with eLISA: it would either contribute 1.5 billion yuan directly or develop its own scaled-down, 8-billion-yuan version of eLISA that would coordinate closely with the European effort, sharing data. Heinzel recommends that a united Chinese group work on one of these less ambitious options.

The direct contribution from China in particular could be a boon for eLISA. Originally, NASA

collaborated with ESA on a planned space-based gravitational-wave observatory, named LISA. But the United States pulled out of LISA five years ago and ESA had to pare down the mission, resulting in the eLISA proposal. China's entry into the project could fill that hole, says Rainer Weiss, a physicist at the Massachusetts Institute of Technology in Cambridge, who is credited as the chief inventor of LIGO. This would perhaps allow Europe to pursue a design closer to that of LISA, which was better equipped than the eLISA proposal and would have had a longer mission lifetime.

A decision is needed soon if China is to achieve a launch date around 2030, cautions Heinzel. "Now is the time to do very serious technology development," he says. "It is time to start making decisions."

Nature **531**, 150–151 (10 March 2016) doi:10.1038/531150a

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

Related stories and links

From nature.com

- **Successful test drive for space-based gravitational-wave detector**
25 February 2016
- **Einstein's gravitational waves found at last**
11 February 2016
- **Freefall space cubes are test for gravitational wave spotter**
17 November 2015

From elsewhere

- [eLISA](#)
- [TianQin](#)

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

4 comments

[Subscribe to comments](#)



Gatot Soedarto • 2016-03-09 10:29 PM

Misled by Einstein's theory of relativity. Gravity is force, and it come due the effects of the well balance of universe. Where does gravity come from?

<http://oejicoba.blogspot.co.id/2016/03/where-does-gravity-come-from.html> Read more in this site: End of the 9 biggest unsolved mysteries in physics,etc.



Xinhang Shen • 2016-03-09 08:19 PM

Gravitation waves may exist, but GTR is wrong because the relativistic spacetime model has been disproved theoretically http://www.eurekalert.org/pub_releases/2016-03/ngpi-tst030116.php.



Gatot Soedarto • 2016-03-10 12:39 AM

Contradiction between STR and GTR: Einstein's general relativity resolved conflicts between Newton theory of gravity and the special theory of relativity, but make a big conflict with his special relativity itself about the existing of aether. Special relativity : aether do not exist, in this theory Einstein believe in Michelson-Morley experiments. General relativity: aether do exist, in this theory Einstein didn't believe in Michelson-Morley experiments. GTR: Light have no mass, can't bent by a gravitational field, but that light is bent by refraction. Einstein make a fatal mistake for his ignorance of the light refraction. Read more in relation with black hole and gravitational lens- <http://oejicoba.blogspot.co.id/2016/03/black-hole-is-made-based-on-unscientific.html>



Pentcho Valev • 2016-03-09 07:39 AM

"The LIGO discovery could increase their chances of success." Not very likely. The LIGO system is fraudulent by design: <http://beforeitsnews.com/space/2016/02/gravitational-wave-discovery-fraud-or-real-scientists-leave-device-running-unattended-as-they-head-to-hotel-2496320.html> "The LIGO team includes a small group of people whose job is to create blind injections—bogus evidence of a gravitational wave—as a way of keeping the scientists on their toes. Although everyone knew who the four people in that group were, "we didn't know what, when, or whether," Gabriela González, the collaboration's spokeswoman, said. During Initial LIGO's final run, in 2010, the detectors picked up what appeared to be a strong signal. The scientists analyzed it intensively for six months, concluding that it was a gravitational wave from somewhere in the constellation of Canis Major. Just before they submitted their results for publication, however, they learned that the signal was a fake. This time through, the blind-injection group swore that they had nothing to do with the signal. Marco Drago thought that their denials might also be part of the test, but Reitze, himself a member of the quartet, had a different concern. "My worry was—and you can file this under the fact that we are just paranoid cautious about making a false claim—could somebody have done this maliciously?" he said. "Could somebody have somehow faked a signal in our detector that we didn't know about?" Reitze, Weiss, González, and a handful of others considered who, if anyone, was familiar enough with both the apparatus and the algorithms to have spoofed the system and covered his or her tracks. There were only four candidates, and none of them had a plausible motive." The Nobel prize was not a motive? The 2010 event was a dress rehearsal; the premiere took place recently:

<http://motls.blogspot.bg/2016/02/ligo-journal-servers-behind-scenes.html> Luboš Motl: " On

September 9th, the LIGO folks were already convinced that they would discover the waves soon. Some of them were thinking what they would buy for the Nobel prize and all of them had to make an online vote about the journal where the discovery should be published. It has to be Physical Review Letters because PRL (published by the APS) is the best journal for the Nobel-prize-caliber papers, the LIGO members decided. Five days later, Advanced LIGO made the discovery. Four more days later, as you know, they officially started Advanced LIGO. ;-) "

Pentcho Valev

See other News & Comment articles from *Nature*

Nature ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER

NATURE | NEWS

Statisticians issue warning over misuse of *P* values

Policy statement aims to halt missteps in the quest for certainty.

Monya Baker

07 March 2016

Misuse of the *P* value — a common test for judging the strength of scientific evidence — is contributing to the number of research findings that cannot be reproduced, the American Statistical Association (ASA) warns in a statement released today¹. The group has taken the unusual step of issuing principles to guide use of the *P* value, which it says cannot determine whether a hypothesis is true or whether results are important.

This is the first time that the 177-year-old ASA has made explicit recommendations on such a foundational matter in statistics, says executive director Ron Wasserstein. The society's members had become increasingly concerned that the *P* value was being misapplied in ways that cast doubt on statistics generally, he adds.

In its statement, the ASA advises researchers to avoid drawing scientific conclusions or making policy decisions based on *P* values alone. Researchers should describe not only the data analyses that produced statistically significant results, the society says, but all statistical tests and choices made in calculations. Otherwise, results may seem falsely robust.

Véronique Kiermer, executive editor of the Public Library of Science journals, says that the ASA's statement lends weight and visibility to longstanding concerns over undue reliance on the *P* value. "It is also very important in that it shows statisticians, as a profession, engaging with the problems in the literature outside of their field," she adds.



How scientists fool themselves – and how they can stop

Weighing the evidence

P values are commonly used to test (and dismiss) a 'null hypothesis', which generally states that there is no difference between two groups, or that there is no correlation between a pair of characteristics. The smaller the *P* value, the less likely an observed set of values would occur by chance — assuming that the null hypothesis is true. A *P* value of 0.05 or less is generally taken to mean that a finding is statistically significant and warrants publication. But that is not necessarily true, the ASA statement notes.

A *P* value of 0.05 does not mean that there is a 95% chance that a given hypothesis is correct. Instead, it signifies that if the null hypothesis is true, and all other assumptions made are valid, there is a 5% chance of obtaining a result at least as extreme as the one observed. And a *P* value cannot indicate the

importance of a finding; for instance, a drug can have a statistically significant effect on patients' blood glucose levels without having a therapeutic effect.



Giovanni Parmigiani, a biostatistician at the Dana Farber Cancer Institute in Boston, Massachusetts, says that misunderstandings about what information a *P* value provides often crop up in textbooks and practice manuals. A course correction is long overdue, he adds. "Surely if this happened twenty years ago, biomedical research could be in a better place now."

**Scientific method:
Statistical errors**

Frustration abounds

Criticism of the *P* value is nothing new. In 2011, researchers trying to raise awareness about false positives gamed an analysis to reach a statistically significant finding: that listening to music by the Beatles makes undergraduates younger². More controversially, in 2015, a set of documentary filmmakers published conclusions from a purposely shoddy clinical trial — supported by a robust *P* value — to show that eating chocolate helps people to lose weight. (The article has since been retracted.)

But Simine Vazire, a psychologist at the University of California, Davis, and editor of the journal *Social Psychological and Personality Science*, thinks that the ASA statement could help to convince authors to disclose all of the statistical analyses that they run. "To the extent that people might be sceptical, it helps to have statisticians saying, 'No, you can't interpret *P* values without this information,' she says.



**Statistics: *P* values
are just the tip of the
iceberg**

More drastic steps, such as the ban on publishing papers that contain *P* values instituted by at least one journal, could be counter-productive, says Andrew Vickers, a biostatistician at Memorial Sloan Kettering Cancer Center in New York City. He compares attempts to bar the use of *P* values to addressing the risk of automobile accidents by warning people not to drive — a message that many in the target audience would probably ignore. Instead, Vickers says that researchers should be instructed to "treat statistics as a science, and not a recipe".

But a better understanding of the *P* value will not take away the human impulse to use statistics to create an impossible level of confidence, warns Andrew Gelman, a statistician at Columbia University in New York City.

"People want something that they can't really get," he says. "They want certainty."

Nature 531, 151 (10 March 2016) doi:10.1038/nature.2016.19503

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

References

1. Wasserstein, R. L. & Lazar, N. A. advance online publication *The American Statistician* (2016).

Show context

2. Simmons, J. P., Nelson, L. D. & Simonsohn, U. *Psychol. Sci.* **22**, 1359–1366 (2011).

Show context

[Article](#) [PubMed](#)

Related stories and links

From nature.com

- **Smart software spots statistical errors in psychology papers**

28 October 2015

- **How scientists fool themselves – and how they can stop**

07 October 2015

- **Statistics: P values are just the tip of the iceberg**

28 April 2015

- **Psychology journal bans P values**

26 February 2015

- **Number crunch**

12 February 2014

- **Scientific method: Statistical errors**

12 February 2014

- **Weak statistical standards implicated in scientific irreproducibility**

11 November 2013

- **Nature special: Challenges in irreproducible research**

From elsewhere

- **American Statistical Association**
-

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

Comments

[Subscribe to comments](#)

There are currently no comments.

See other News & Comment articles from *Nature*

Nature ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER

NATURE | NEWS

Five million US seeds banked for resurrection experiment

Project Baseline will monitor effects of climate change on plant evolution.

Daniel Cressey

08 March 2016



Print

Michael Marquand/Getty

The Joshua tree (*Yucca brevifolia*) is one of dozens of species in the Project Baseline seed bank.

In a vault kept at -18°C in Fort Collins, Colorado, more than 5 million seeds now lie frozen in time — destined to wait for up to 50 years until evolutionary scientists earn permission to experiment with them.

Unlike most seed banks, which aim to protect biological diversity, Project Baseline is designed to enable precise, controlled studies of how plants are evolving in response to climate change and environmental degradation. Taken from around 250 locations across the continental United States and stored at a US Department of Agriculture facility, the seeds represent some 60 species.

Scientists began collecting the seeds in earnest in 2012, backed by a US\$1.3-million grant from the US National Science Foundation (NSF). They took care to gather specimens in a wide variety of environments and to cover a multitude of plant types, from the humble radish (*Raphanus sativus*) to the iconic Joshua tree (*Yucca brevifolia*).

The collection phase is now complete, says project lead investigator Julie Etterson, a plant biologist at the University of Minnesota Duluth. Earlier this year, she and her colleagues published a paper in the *American Journal of Botany* introducing Project Baseline to the community (J. R. Etterson *et al.* *Am. J. Bot.* **103**, 164–173; 2016).

To find out whether species are evolving in response to human pressures such as climate change, scientists have previously observed differences in similar species living at various sites or studied one site over time, charting how plants change along with the site. But it can be difficult to distinguish between changes that are the result of evolution — the selection of traits over generations owing to the survival of certain individuals — and those that are due to the ability of individual plants to react to a changing environment, called plasticity.

Related stories

- Turning point: Heather Schneider
- Wild flower blooms again after 30,000 years on ice
- Seed banks susceptible to sham samples

More related stories

Project Baseline will allow scientists to grow stored seeds side by side with those from plants that were left to evolve, in identical conditions: any differences can then be attributed to evolution.

"I think it's terrific," says Richard Lenski, who studies evolution in bacteria at Michigan State University in East Lansing. "To some extent, museum specimens and even natural seed banks allow scientists to make these comparisons today, but not in the in-depth, systematic and well-thought-out way that this project will allow."

Questions that could be explored include whether the early flowering observed in some plants in conjunction with global warming is attributable to evolution or plasticity, and how rates of evolution vary between different populations of the same species. Genetic sequencing will help researchers to discover which genes are linked to traits that have been selected for. It could also test predictions, such as that low genetic variation increases extinction rates, and that evolution occurs through many small genetic changes rather than a few large ones. "The list of hypotheses is really only limited by the imagination," says Etterson.

Back to life

Project Baseline breathes new life into a field known as resurrection ecology. Its best-known experiments hatched invertebrate eggs that had been naturally preserved in lake sediments, and compared the offspring with those of recently laid eggs. A now-classic example, from the lab of environmental scientist Nelson Hairston at Cornell University in Ithaca, New York, used sediments from Lake Constance in central Europe to prove that water fleas (*Daphnia galeata*) had rapidly evolved tolerance for toxic cyanobacteria (N. G. Hairston *et al.* *Nature* **401**, 446; 1999).

Because Project Baseline actively lays the foundation for future research, rather than relying on what nature has sequestered in the past, it is a "kind of visionary project", says Hairston.

It does assume that there will be observable environmental changes at the sites from which the seeds were collected, notes Charles Kerfoot, a biologist at Michigan Technological University in Houghton and another pioneer in resurrection ecology. But such differences are guaranteed because of climate change, he says: "This is a group that's not in denial."

Exactly when the scientists will wake the seeds in the vault from stasis is less clear. Project Baseline's first call for proposals to work with the specimens is planned for 2018, and Etterson says that the first seeds could be planted as soon as 2020. She hopes to get at least one use out of the project herself before she retires.

The timescales are long compared with both the average evolution study and the average NSF grant, say researchers, but that makes Project Baseline special. "This is really different," says Samuel Scheiner, the director of the NSF programme that funded the project, "but exactly what we need to do if we're going to study global change."

Nature **531**, 152 (10 March 2016) doi:10.1038/531152a

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

Related stories and links

From nature.com

- **Turning point: Heather Schneider**
29 July 2015
- **Wild flower blooms again after 30,000 years on ice**
21 February 2012
- **Seed banks susceptible to sham samples**
11 March 2011

From elsewhere

- **Project Baseline**

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be redirected back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

Comments

[Subscribe to comments](#)

There are currently no comments.

[See other News & Comment articles from *Nature*](#)

Nature ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER

NATURE | NEWS

First Zika-linked birth defects detected in Colombia

Cases may signal start of anticipated wave of birth defects in country hit hard by Zika virus.

Declan Butler

04 March 2016



Luis Robayo/AFP/Getty Images

Brochures with information about the Zika virus have been delivered to pregnant women in Colombia.

Researchers have found Colombia's first cases of birth defects linked to the Zika virus, *Nature* has learned — which are likely forerunners of a widely anticipated wave of Zika-related birth defects in the country.

The discovery is perhaps no surprise: the virus arrived in Colombia last September, and the country is second only to Brazil in terms of the number of people infected with Zika.

But Colombian researchers hope that plans put in place to closely monitor pregnant women can help to better establish the magnitude of the threat posed to fetuses by Zika. That is a crucial question that scientists have not so far been able to answer with the data from Brazil.

Researchers have diagnosed one newborn with microcephaly — an abnormally small head — and two

others with congenital brain abnormalities, says Alfonso Rodriguez-Morales, who chairs the Colombian Collaborative Network on Zika (RECOLZIKA), which made the diagnoses. All three tested positive for the presence of Zika virus. The researchers have submitted a report of their detections to a scientific journal.



Zika researchers release real-time data on viral infection study in monkeys

Rodriguez-Morales, an infectious-diseases epidemiologist at the Technological University of Pereira in western Colombia, says that he expects to see a rise in cases of Zika-linked birth defects starting in two or three months' time. The RECOLZIKA group — a network of researchers and public-health institutions across Colombia — are already investigating a handful of other suspected cases of microcephaly, which have a possible link to Zika.

The next wave?

Brazil is the only country so far to report a large surge in newborns with microcephaly that coincides with outbreaks of Zika virus. By the time the alarm over a possible microcephaly link was raised there (in October 2015), Zika infections had already peaked in many parts of the country, because the virus first reached Brazil at the beginning of last year.

In Colombia, by contrast, researchers detected the first Zika cases in September, and by December had set up national tracking programmes to monitor pregnant women for signs of infection, and to spot early signs of birth defects in fetuses. Since then, researchers have been waiting attentively to see whether their country might experience a similar rise in birth defects.

The true size of Brazil's surge in microcephaly cases is unknown. The country's health ministry says that 5,909 suspected microcephaly cases have been registered since early November, but only 1,687 of them have been investigated so far. Of those, 1,046 have been discarded as false positives, and 641 have been confirmed. (A link with the Zika virus has been confirmed by molecular-lab tests in 82 cases.)

Given that Brazil reported only 147 cases of microcephaly in 2014, the reported increase in cases since November suggests a marked rise in the number of babies born with the condition. But the 2014 figure is a "huge underestimate", says Lavinia Schüler-Faccini, a geneticist who specialises in birth defects at the Federal University of Rio Grande do Sul, Brazil, and president of the Brazilian Society of Genetic Medicine. She says that according to the frequency of microcephaly typically observed in regions around the world, one would expect to see 300–600 cases of severe microcephaly in any given year in Brazil, and around 1,500 less-severe ones.



Proving Zika link to birth defects poses huge challenge

The search for cases of microcephaly in Brazil since October is probably turning up many mild cases that previously went unnoticed — so that the reported surge looks higher than it really is. Still, Schüler-Faccini and other clinicians say there is a real problem. They have observed first-hand a marked

increase in the number of unusually severe cases of microcephaly, they say.

To be prepared to better interpret any imminent peak in birth defects in Colombia, RECOLZIKA plans to look at historical cases to establish a baseline for the annual numbers of birth defects in different regions. It is also setting up a study to analyse patterns in the distribution of head-circumference measurements recorded in obstetrics units regionally throughout the country, to get a better idea of the local range of normal values.



Luis Robayo/AFP/Getty Images

A pregnant woman holds a mosquito net — delivered by Colombia's health ministry to ward off Zika virus infection — in Santiago de Cali, Colombia.

Zika's link to microcephaly

It has also not been possible so far from Brazilian data to quantify the extent to which Zika virus is linked to the rise in microcephaly. The latest data from Brazil's ministry of health show that increased cases of microcephaly and/or congenital malformations of the central nervous system are still concentrated in the northeast — raising questions as to whether other factors, perhaps specific to this region, might also be in play.

Clinical evidence leaves little doubt that a link between Zika and microcephaly exists: the virus has been detected in amniotic fluid, in the cerebrospinal fluid of affected babies and in the brains of stillborn fetuses and those aborted after the detection of severe malformations during pregnancy.

But there are also many other possible causes of microcephaly, including a group of infections that are collectively called STORCH (syphilis, toxoplasmosis, other infections, rubella, cytomegalovirus infection and herpes simplex), which are known to cause birth defects. Exposure to toxic chemicals and the consumption of alcohol during pregnancy can also cause the condition.



**Spectre of Ebola
haunts Zika response**

"There is a clear need for a full assessment of other detailed causes of microcephaly, such as STORCH, and even non-infectious causes," says Rodriguez-Morales. Brazil's health ministry has stated that it is carrying out tests for such causes, but it has not made public how many of the confirmed microcephaly cases are attributable to these.

Healthy comparisons

A key question in assessing the scale of the threat that Zika may pose to fetuses is how many pregnant women infected with Zika — in particular during the first trimester, when the fetus is most vulnerable — nonetheless give birth to healthy babies. RECOLZIKA researchers hope to help to answer this through their monitoring programme.

The risk posed by Zika may well be lower than that of other diseases that are known to cause microcephaly such as toxoplasmosis and rubella, says Rodriguez-Morales. That is a preliminary estimate, he says, based on back-of-the-envelope calculations of the reported numbers of confirmed cases of microcephaly and congenital disorders, compared to the number of pregnant women in regions experiencing Zika epidemics.

But even if its risk does turn out to be low, Zika could still lead to many cases because a large number of pregnant women in the Americas are likely to become infected with the virus.



**Maternal health:
Ebola's lasting legacy**

The biggest risk to pregnant women is right now, rather than in the long term. The epidemic is sweeping so quickly through the Americas that much of the population, including young women, will become naturally vaccinated by their exposure to the virus. As population immunity increases, the Zika epidemic is likely to fade quickly, and it will become endemic with only occasional flare ups.

In a modelling study posted to the preprint server bioRxiv¹ on 29 February, US researchers noted that the risk of prenatal Zika virus exposure "should decrease dramatically following the initial wave of disease, reaching almost undetectable levels".

Nature **531**, 153 (10 March 2016) doi:10.1038/nature.2016.19502

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

References

1. Bewick, S., Fagan, W. F., Calabrese, J. M. & Agusto, F. Preprint at bioRxiv
<http://dx.doi.org/10.1101/041897> (2016).

Show context

Related stories and links

From nature.com

- **Spectre of Ebola haunts Zika response**
02 March 2016
- **Zika researchers release real-time data on viral infection study in monkeys**
23 February 2016
- **Proving Zika link to birth defects poses huge challenge**
09 February 2016
- **The next steps on Zika**
02 February 2016
- **Zika virus: Brazil's surge in small-headed babies questioned by report**
28 January 2016

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

Comments

[Subscribe to comments](#)

There are currently no comments.

See other News & Comment articles from *Nature*

Nature ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER

Just under a year ago, a molecular-biology technique was thrust onto the world stage. Researchers in China announced that they had used the nascent gene-editing tool CRISPR–Cas9 to modify the genomes of human embryos, triggering a major ethics debate.

Yet while this controversy has been playing out, researchers the world over have rushed to use the tool to tinker with the genomes of other human cells, viruses, bacteria, animals and plants, and it's in these contexts that the technique promises to have more immediate impact. This issue of *Nature* examines what's going on at the CRISPR frontiers.

Biologists are using CRISPR–Cas9 to better understand genomes — not just by editing DNA, but by devising variations on the technique to precisely manipulate the activity of genes (see page 156). And, armed for the first time with a method that can easily introduce genetic changes to many animals, researchers have edited a veritable menagerie of beasts — from ferrets to elephants to koi carp (see page 160) — in an attempt to combat disease, improve agriculture and even make designer pets.

Such advances in gene editing are creating

upheaval for the regulatory bodies that are responsible for approving genetically engineered products — it's a "powder keg waiting to explode", writes Jennifer Kuzma, a science-policy researcher at North Carolina State University in Raleigh, on page 165. She calls for more openness and honesty than has characterized past discussions of biotechnology, and for a regulatory system that better factors in societal views as well as science.

CRISPR–Cas9 may be democratizing gene editing in the laboratory, but Todd Kuiken, who studies science policy at the Wilson Center, a think tank in Washington DC, argues on page 167 that the revolution has not yet swept into home workshops or citizen-science community spaces. Contrary to reports in the popular media, he says, few CRISPR creations are likely to come from the labs of do-it-yourself biologists any time soon. However, this group is arguably ahead of the scientific establishment when it comes to thinking about how to use the technology safely.

For better or for worse, CRISPR–Cas9 is transforming biology. We are now at the dawn of the gene-editing age. ■



EVERYWHERE

A special issue explores what it means to be living in an age of gene editing.

ILLUSTRATION BY CHRIS LABROOY



CRISPR EVERYWHERE

A *Nature* special issue

nature.com/crispr



RIDING THE CRISPR WAVE

Biologists are embracing the power of gene-editing tools to explore genomes.

BY HEIDI LEDFORD

Whenever a paper about CRISPR–Cas9 hits the press, the staff at Addgene quickly find out. The non-profit company is where study authors often deposit molecular tools that they used in their work, and where other scientists immediately turn to get them. “We get calls within minutes of a hot paper publishing,” says Joanne Kamens, executive director of the company in Cambridge, Massachusetts.

Addgene’s phones have been ringing a lot since early 2013, when researchers first reported^{1–3} that they had used the CRISPR–Cas9 system to slice the genome in human cells at sites of their choosing. “It was all hands on deck,” Kamens says. Since then, molecular biologists have rushed to adopt the technique,



CRISPR EVERYWHERE

A *Nature* special issue
nature.com/crispr

really an extraordinarily powerful way to understand how the genome works," says Daniel Bauer, a haematologist at the Boston Children's Hospital in Massachusetts. "It's really opened the number of questions you can address," adds Peggy Farnham, a molecular biologist at the University of Southern California, Los Angeles. "It's just so fun."

Here, *Nature* examines five ways in which CRISPR–Cas9 is changing how biologists can tinker with cells.

BROKEN SCISSORS

There are two chief ingredients in the CRISPR–Cas9 system: a Cas9 enzyme that snips through DNA like a pair of molecular scissors, and a small RNA molecule that directs the scissors to a specific sequence of DNA to make the cut. The cell's native DNA repair machinery generally mends the cut — but often makes mistakes.

That alone is a boon to scientists who want to disrupt a gene to learn about what it does. The genetic code is merciless: a minor error introduced during repair can completely alter the sequence of the protein it encodes, or halt its production altogether. As a result, scientists can study what happens to cells or organisms when the protein or gene is hobbled.

But there is also a different repair pathway that sometimes mends the cut according to a DNA template. If researchers provide the template, they can edit the genome with nearly any sequence they desire at nearly any site of their choosing.

In 2012, as laboratories were racing to demonstrate how well these gene-editing tools could cut human DNA, one team decided to take a different approach. "The first thing we did: we broke the scissors," says Jonathan Weissman, a systems biologist at the University of California, San Francisco (UCSF).

Weissman learned about the approach from Stanley Qi, a synthetic biologist now at Stanford University in California, who mutated the Cas9 enzyme so that it still bound DNA at the site that matched its guide RNA, but no longer sliced it. Instead, the enzyme stalled there and blocked other proteins from transcribing that DNA into RNA. The hacked system allowed them to turn a gene off, but without altering the DNA sequence⁴.

The team then took its 'dead' Cas9 and tried something new: the researchers tethered it to part of another protein, one that activates gene expression. With a few other tweaks, they had built a way to turn genes on and off at will⁵.

Several labs have since published variations on this method; many more are racing to harness it for their research⁶ (see 'Hacking CRISPR'). One popular application is to rapidly generate hundreds of different cell lines, each containing a different guide RNA that targets a particular gene. Martin Kampmann, another systems biologist at UCSF, hopes to screen such cells to learn whether flipping certain genes on or off affects the survival of neurons exposed to

toxic protein aggregates — a mechanism that is thought to underlie several neurodegenerative conditions, including Alzheimer's disease. Kampmann had been carrying out a similar screen with RNA interference (RNAi), a technique that also silences genes and can process lots of molecules at once, but which has its drawbacks. "RNAi is a shotgun with well-known off-target effects," he says. "CRISPR is the scalpel that allows you to be more specific."

Weissman and his colleagues, including UCSF systems biologist Wendell Lim, further tweaked the method so that it relied on a longer guide RNA, with motifs that bound to different proteins. This allowed them to activate or inhibit genes at three different sites all in one experiment⁷. Lim thinks that the system can handle up to five operations at once. The limit, he says, may be in how many guide RNAs and proteins can be stuffed into a cell. "Ultimately, it's about payload."

That combinatorial power has drawn Ron Weiss, a synthetic biologist at the Massachusetts Institute of Technology (MIT) in Cambridge, into the CRISPR–Cas9 frenzy. Weiss and his colleagues have also created multiple gene tweaks in a single experiment⁸, making it faster and easier to build complicated biological circuits that could, for example, convert a cell's metabolic machinery into a biofuel factory. "The most important goal of synthetic biology is to be able to program complex behaviour via the creation of these sophisticated circuits," he says.

CRISPR EPIGENETICS

When geneticist Marianne Rots began her career, she wanted to unearth new medical cures. She studied gene therapy, which targets genes mutated in disease. But after a few years, she decided to change tack. "I reasoned that many more diseases are due to disturbed gene-expression profiles, not so much the single genetic mutations I had been focused on," says Rots, at the University Medical Center Groningen in the Netherlands. The best way to control gene activity, she thought, was to adjust the epigenome, rather than the genome itself.

The epigenome is the constellation of chemical compounds tacked onto DNA and the DNA-packaging proteins called histones. These can govern access to DNA, opening it up or closing it off to the proteins needed for gene expression. The marks change over time: they are added and removed as an organism develops and its environment shifts.

In the past few years, millions of dollars have been poured into cataloguing these epigenetic marks in different human cells, and their patterns have been correlated with everything from brain activity to tumour growth. But without the ability to alter the marks at specific sites, researchers are unable to determine whether they cause biological changes. "The field has met a lot of resistance because we haven't had the kinds of tools that geneticists have had, where they can go in and directly test the function of

which can be used to alter the genome of almost any organism with unprecedented ease and finesse. Addgene has sent 60,000 CRISPR-related molecular tools — about 17% of its total shipments — to researchers in 83 countries, and the company's CRISPR-related pages were viewed more than one million times in 2015.

Much of the conversation about CRISPR–Cas9 has revolved around its potential for treating disease or editing the genes of human embryos, but researchers say that the real revolution right now is in the lab. What CRISPR offers, and biologists desire, is specificity: the ability to target and study particular DNA sequences in the vast expanse of a genome. And editing DNA is just one trick that it can be used for. Scientists are hacking the tools so that they can send proteins to precise DNA targets to toggle genes on or off, and even engineer entire biological circuits — with the long-term goal of understanding cellular systems and disease.

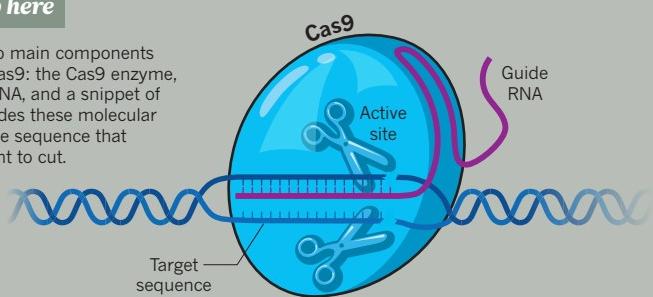
"For the humble molecular biologist, it's

HACKING CRISPR

By modifying the molecular machinery that powers CRISPR–Cas9 gene editing, scientists can probe the functions of genes and gene regulators with unprecedented specificity.

Snip snip here

There are two main components of CRISPR–Cas9: the Cas9 enzyme, which cuts DNA, and a snippet of RNA that guides these molecular scissors to the sequence that scientists want to cut.

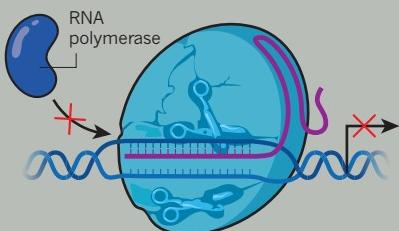


Broken scissors

The Cas9 enzyme can be broken so that it no longer cuts DNA. But with the right guide RNA, it can still attach to specific parts of the genome.

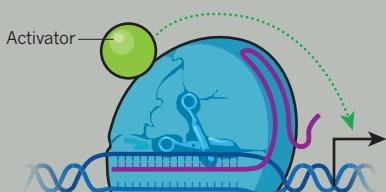
CRISPR inhibition

A broken, or ‘dead’, Cas9 enzyme will block the binding of other proteins, such as RNA polymerase, needed to express a gene.



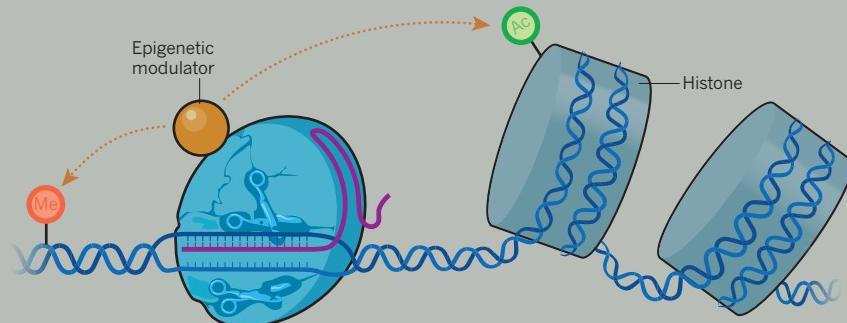
CRISPR activation

An activating protein can be attached to a dead Cas9 protein to stimulate expression of a specific gene.



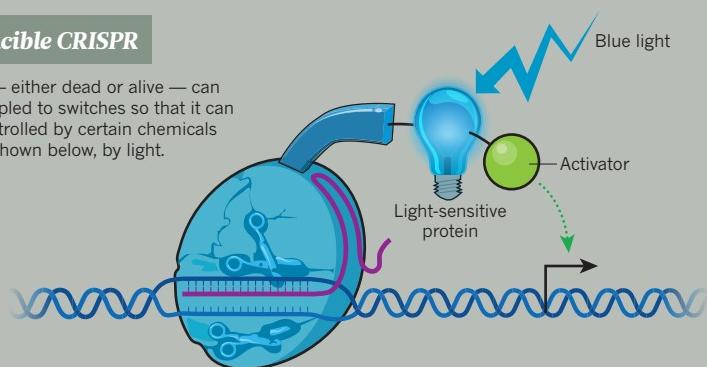
CRISPR epigenetics

A broken Cas9 enzyme can be coupled to epigenetic modifiers, such as those that add methyl groups (Me) to DNA or acetyl groups (Ac) to histone proteins. This will allow researchers to study how precisely placed modifications affect gene expression and DNA dynamics.



Inducible CRISPR

Cas9 — either dead or alive — can be coupled to switches so that it can be controlled by certain chemicals or, as shown below, by light.



a gene,” says Jeremy Day, a neuroscientist at the University of Alabama at Birmingham.

CRISPR–Cas9 could turn things around. In April 2015, Charles Gersbach, a bioengineer at Duke University in Durham, North Carolina, and his colleagues published⁵ a system for adding acetyl groups — one type of epigenetic mark — to histones using the broken scissors to carry enzymes to specific spots in the genome.

The team found that adding acetyl groups to proteins that associate with DNA was enough to send the expression of targeted genes soaring, confirming that the system worked and that, at this location, the epigenetic marks had an effect. When he published the work, Gersbach deposited his enzyme with Addgene so that other research groups could use it — and they quickly did. Gersbach predicts that a wave of upcoming papers will show a synergistic effect when multiple epigenetic markers are manipulated at once.

The tools need to be refined. Dozens of enzymes can create or erase an epigenetic mark on DNA, and not all of them have been amenable to the broken-scissors approach. “It turned out to be harder than a lot of people were expecting,” says Gersbach. “You attach a lot of things to a dead Cas9 and they don’t happen to work.” Sometimes it is difficult to work out whether an unexpected result arose because a method did not work well, or because the epigenetic mark simply doesn’t matter in that particular cell or environment.

Rots has explored the function of epigenetic marks on cancer-related genes using older editing tools called zinc-finger proteins, and is now adopting CRISPR–Cas9. The new tools have democratized the field, she says, and that has already had a broad impact. People used to say that the correlations were coincidental, Rots says — that if you rewrite the epigenetics it will have no effect on gene expression. “But now that it’s not that difficult to test, a lot of people are joining the field.”

CRISPR CODE CRACKING

Epigenetic marks on DNA are not the only genomic code that is yet to be broken. More than 98% of the human genome does not code for proteins. But researchers think that a fair chunk of this DNA is doing something important, and they are adopting CRISPR–Cas9 to work out what that is.

Some of it codes for RNA molecules — such as microRNAs and long non-coding RNAs — that are thought to have functions apart from making proteins. Other sequences are ‘enhancers’ that amplify the expression of the genes under their command. Most of the DNA sequences linked to the risk of common diseases lie in regions of the genome that contain non-coding RNA and enhancers. But before CRISPR–Cas9, it was difficult for researchers to work out what those sequences do. “We didn’t have a good way to functionally annotate the non-coding genome,” says Bauer. “Now our

experiments are much more sophisticated."

Farnham and her colleagues are using CRISPR–Cas9 to delete enhancer regions that are found to be mutated in genomic studies of prostate and colon cancer. The results have sometimes surprised her. In one unpublished experiment, her team deleted an enhancer that was thought to be important, yet no gene within one million bases of it changed expression. "How we normally classify the strength of a regulatory element is not corresponding with what happens when you delete that element," she says.

More surprises may be in store as researchers harness CRISPR–Cas9 to probe large stretches of regulatory DNA. Groups led by geneticists David Gifford at MIT and Richard Sherwood at the Brigham and Women's Hospital in Boston used the technique to create mutations across a 40,000-letter sequence, and then examined whether each change had an effect on the activity of a nearby gene that made a fluorescent protein¹⁰. The result was a map of DNA sequences that enhanced gene expression, including several that had not been predicted on the basis of gene regulatory features such as chromatin modifications.

Delving into this dark matter has its challenges, even with CRISPR–Cas9. The Cas9 enzyme will cut where the guide RNA tells it to, but only if a specific but common DNA sequence is present near the cut site. This poses little difficulty for researchers who want to silence a gene, because the key sequences almost always exist somewhere within it. But for those who want to make very specific changes to short, non-coding RNAs, the options can be limited. "We cannot take just any sequence," says Reuven Agami, a researcher at the Netherlands Cancer Institute in Amsterdam.

Researchers are scouring the bacterial kingdom for relatives of the Cas9 enzyme that recognize different sequences. Last year, the lab of Feng Zhang, a bioengineer at the Broad Institute of MIT and Harvard in Cambridge, characterized a family of enzymes called Cpf1 that work similarly to Cas9 and could expand sequence options¹¹. But Agami notes that few alternative enzymes found so far work as well as the most popular Cas9. In the future, he hopes to have a whole collection of enzymes that can be targeted to any site in the genome. "We're not there yet," he says.

CRISPR SEES THE LIGHT

Gersbach's lab is using gene-editing tools as part of an effort to understand cell fate and how to manipulate it: the team hopes one day to grow tissues in a dish for drug screening and cell therapies. But CRISPR–Cas9's effects are permanent, and Gersbach's team needed to turn genes on and off transiently, and in very specific locations in the tissue. "Patterning a blood vessel demands a high degree of control," he says.

Gersbach and his colleagues took their broken, modified scissors — the Cas9 that

could now activate genes — and added proteins that are activated by blue light. The resulting system triggers gene expression when cells are exposed to the light, and stops it when the light is flicked off¹². A group led by chemical biologist Moritoshi Sato of the University of Tokyo rigged a similar system¹³, and also made an active Cas9 that edited the genome only after it was hit with blue light¹⁴.

Others have achieved similar ends by combining CRISPR with a chemical switch.

the new CRISPR–Cas9 tools to precisely manipulate the genome and epigenome in animal models. "The real power is going to be the integration of those systems," says Dow. This may allow scientists to capture and understand some of the complexity of common human diseases.

Take tumours, which can bear dozens of mutations that potentially contribute to cancer development. "They're probably not all important in terms of modelling a tumour,"

**"I wish I had had this technology sooner.
My postdoc would have been a lot shorter."**

Lukas Dow, a cancer geneticist at Weill Cornell Medical College in New York City, wanted to mutate cancer-related genes in adult mice, to reproduce mutations that have been identified in human colorectal cancers. His team engineered a CRISPR–Cas9 system in which a dose of the compound doxycycline activates Cas9, allowing it to cut its targets¹⁵.

The tools are another step towards gaining fine control over genome editing. Gersbach's team has not patterned its blood vessels just yet: for now, the researchers are working on making their light-inducible system more efficient. "It's a first-generation tool," says Gersbach.

MODEL CRISPR

Cancer researcher Wen Xue spent the first years of his postdoc career making a transgenic mouse that bore a mutation found in some human liver cancers. He slogged away, making the tools necessary for gene targeting, injecting them into embryonic stem cells and then trying to derive mice with the mutation. The cost: a year and US\$20,000. "It was the rate-limiting step in studying disease genes," he says.

A few years later, just as he was about to embark on another transgenic-mouse experiment, his mentor suggested that he give CRISPR–Cas9 a try. This time, Xue just ordered the tools, injected them into single-celled mouse embryos and, a few weeks later — *voilà*. "We had the mouse in one month," says Xue. "I wish I had had this technology sooner. My postdoc would have been a lot shorter."

Researchers who study everything from cancer to neurodegeneration are embracing CRISPR–Cas9 to create animal models of the diseases (see page 160). It lets them engineer more animals, in more complex ways, and in a wider range of species. Xue, who now runs his own lab at the University of Massachusetts Medical School in Worcester, is systematically sifting through data from tumour genomes, using CRISPR–Cas9 to model the mutations in cells grown in culture and in animals.

Researchers are hoping to mix and match

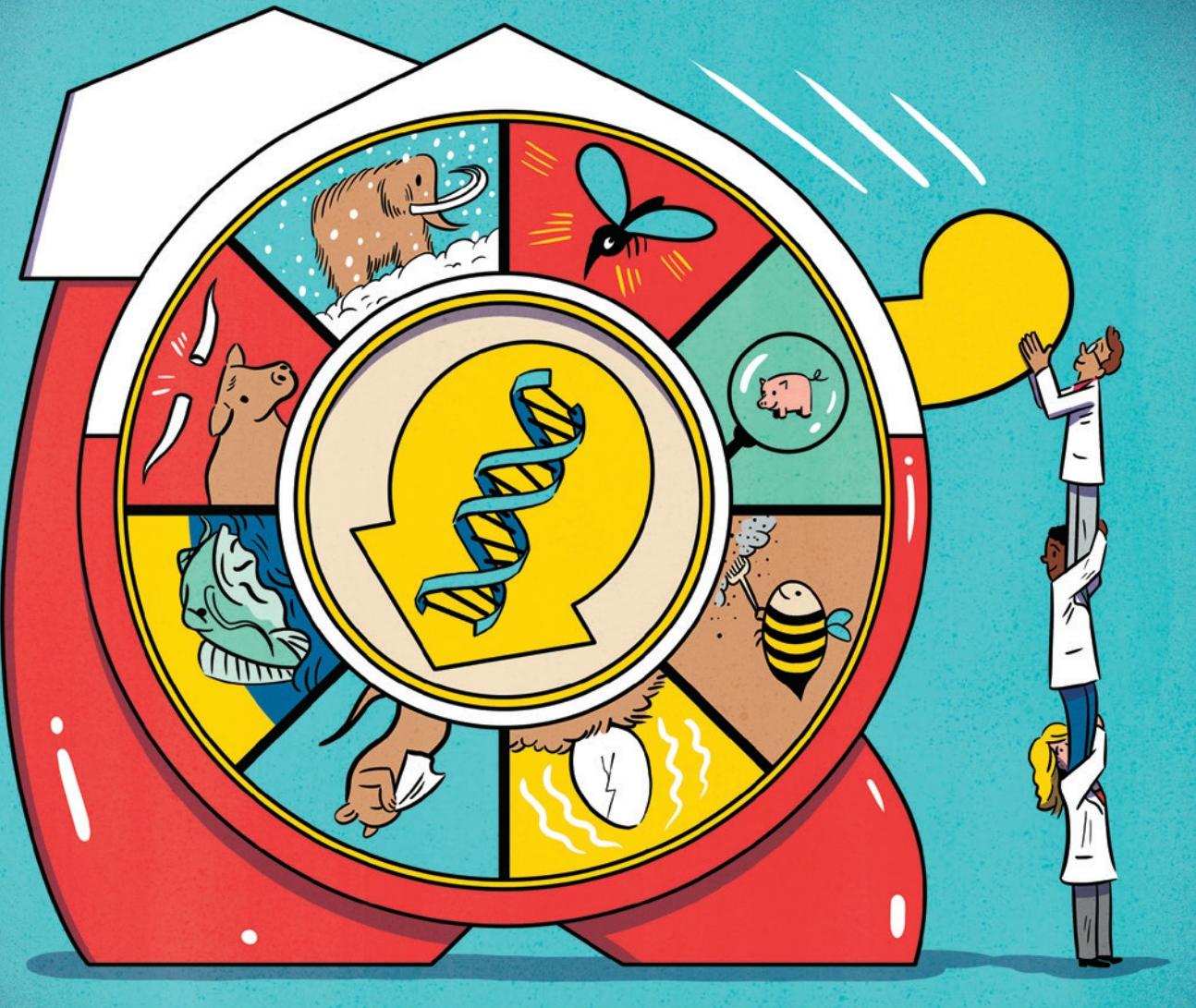
says Dow. "But it's very clear that you're going to need two or three or four mutations to really model aggressive disease and get closer to modelling human cancer." Introducing all of those mutations into a mouse the old-fashioned way would have been costly and time-consuming, he adds.

Bioengineer Patrick Hsu started his lab at the Salk Institute for Biological Studies in La Jolla, California, in 2015; he aims to use gene editing to model neurodegenerative conditions such as Alzheimer's disease and Parkinson's disease in cell cultures and marmoset monkeys. That could recapitulate human behaviours and progression of disease more effectively than mouse models, but would have been unthinkable expensive and slow before CRISPR–Cas9.

Even as he designs experiments to genetically engineer his first CRISPR–Cas9 marmosets, Hsu is aware that this approach may be only a stepping stone to the next. "Technologies come and go. You can't get married to one," he says. "You need to always think about what biological problems need to be solved." ■

Heidi Ledford is a senior reporter for Nature in Cambridge, Massachusetts.

1. Cong, L. et al. *Science* **339**, 819–823 (2013).
2. Mali, P. et al. *Science* **339**, 823–826 (2013).
3. Jinek, M. et al. *eLife* **2**, e00471 (2013).
4. Qi, L. S. et al. *Cell* **152**, 1173–1183 (2013).
5. Gilbert, L. A. et al. *Cell* **159**, 647–661 (2014).
6. Dominguez, A. A., Lim, W. A. & Qi, L. S. *Nature Rev. Mol. Cell. Biol.* **17**, 5–15 (2016).
7. Zalatan, J. G. et al. *Cell* **160**, 339–350 (2015).
8. Kiani, S. et al. *Nature Meth.* **11**, 723–726 (2014).
9. Hilton, I. B. et al. *Nature Biotechnol.* **33**, 510–517 (2015).
10. Rajagopal, N. et al. *Nature Biotechnol.* **34**, 167–174 (2016).
11. Zetsche, B. et al. *Cell* **163**, 759–771 (2015).
12. Polstein, L. R. & Gersbach, C. A. *Nature Chem. Biol.* **11**, 198–200 (2015).
13. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H. & Sato, M. *Chem. Biol.* **22**, 169–174 (2015).
14. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. *Nature Biotechnol.* **33**, 755–760 (2015).
15. Dow, L. E. et al. *Nature Biotechnol.* **33**, 390–394 (2015).



THE CRISPR ZOO

*Birds and bees are just the beginning
for a burgeoning technology.*

BY SARA REARDON

Timothy Doran's 11-year-old daughter is allergic to eggs. And like about 2% of children worldwide who share the condition, she is unable to receive many routine vaccinations because they are produced using chicken eggs.

Doran, a molecular biologist at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Geelong, Australia, thinks that he could solve this problem using the powerful gene-editing tool CRISPR–Cas9. Most egg allergies are caused by one of just four proteins in the white, and when Doran's colleagues altered the gene that encodes one of these in bacteria, the resulting protein no longer triggered a reaction in blood serum from people who were known to be allergic

to it¹. Doran thinks that using CRISPR to edit the gene in chickens could result in hypoallergenic eggs.

The group expects to hatch its first generation of chicks with gene modifications later this year as a proof of concept. Doran realizes that it could be some time before regulators would approve gene-edited eggs, and he hopes that his daughter will have grown out of her allergy by then. "If not, I've got someone ready and waiting to try the first egg," he says.



CRISPR EVERYWHERE

A *Nature* special issue
nature.com/crispr

Chickens are just one of a menagerie of animals that could soon have their genomes reimagined. Until now, researchers had the tools to genetically manipulate only a small selection of animals, and the process was

often inefficient and laborious. With the arrival of CRISPR, they can alter the genes of a wide range of organisms with relative precision and ease. In the past two years alone, the prospect of gene-edited monkeys, mammoths, mosquitoes and more have made headlines as scientists attempt to put CRISPR to use for applications as varied as agriculture, drug production and bringing back lost species. CRISPR-modified animals are even being marketed for sale as pets. "It's allowed us to consider a whole raft of projects we couldn't before," says Bruce Whitelaw, an animal biotechnologist at the Roslin Institute in Edinburgh, UK. "The whole community has wholeheartedly moved towards genome editing."

But regulators are still working out how to deal with such creatures, particularly those intended for food or for release into the wild. Concerns abound about safety and ecological impacts. Even the US director of national intelligence has weighed in, saying that the easy access, low cost and speedy development of genome editing could increase the risk that someone will engineer harmful biological agents.

Eleonore Pauwels, who studies biotechnology regulation at the Wilson Center in Washington DC, says that the burgeoning use of CRISPR in animals offers an opportunity for researchers and policy-makers to engage the public in debate. She hopes that such discussions will help in determining which uses of CRISPR will be most helpful to humans, to other species and to science — and will highlight the limits of the technology. "I think there is a lot of value in humility about how much control we have," she says.

DISEASE CONTROL

Disease resistance is one of the most popular applications for CRISPR in agriculture, and scientists are tinkering across a wide spectrum of animals. Biotechnology entrepreneur Brian Gillis in San Francisco is hoping that the tool can help to stem the dramatic loss of honeybees around the world, which is being caused by factors such as disease and parasites.

Gillis has been studying the genomes of 'hygienic' bees, which obsessively clean their hives and remove sick and infested bee larvae. Their colonies are less likely to succumb to mites, fungi and other pathogens than are those of other strains, and Gillis thinks that if he

can identify genes associated with the behaviour, he might be able to edit them in other breeds to bolster hive health.

But the trait could be difficult to engineer. No hygiene-associated genes have been definitively identified, and the roots of the behaviour may prove complex, says BartJan Fennhout, chairman of Arista Bee Research in Boxmeer, the Netherlands, which studies mite resistance. Moreover, if genes are identified, he says, conventional breeding may be sufficient to confer resistance to new populations, and that might be preferable given the widespread opposition to genetic engineering.

Such concerns don't seem to have slowed down others studying disease resistance. Whitelaw's group at the Roslin Institute is one of several using CRISPR and other gene-editing systems to create pigs that are resistant to viral diseases that cost the agricultural industry hundreds of millions of dollars each year.

Whitelaw's team is using another gene-editing technique to alter immune genes in domestic pigs to match more closely those of warthogs that are naturally resistant to African swine fever, a major agricultural pest². And Randall Prather at the University of Missouri in Columbia has created pigs with a mutated protein on the surface of their cells, which should make them impervious to a deadly respiratory virus³. Other researchers are making cattle that are resistant to the trypanosome parasites that are responsible for sleeping sickness.

Whitelaw hopes that regulators — and sceptical consumers — will be more enthusiastic about animals that have had their genes edited to improve disease resistance than they have been for traits such as growth promotion because of the potential to reduce suffering. And some governments are considering whether CRISPR-modified animals should be regulated in the same way as other genetically modified organisms, because they do not contain DNA from other species.

MAKING DRUGS

Doran's quest to modify allergens in chicken eggs requires delicate control. The trick is to finely adjust a genetic sequence in a way that will stop the protein from triggering an immune reaction in people, but still allow it to perform its normal role in embryonic development. CRISPR has made such precise edits possible for the first time. "CRISPR has been the saviour for trying to tackle allergens," says Mark Tizard, a molecular biologist at CSIRO who works with Doran on chickens.

Using the technique in birds still presents problems. Mammals can be induced to produce extra eggs, which can then be removed, edited, fertilized and replaced. But in birds, the fertilized egg binds closely to the yolk and removing it would destroy the embryo. And because eggs are difficult to access while still inside the hen, CRISPR components cannot be directly injected into the egg itself. By the time the egg is laid, development has proceeded too far for gene editing to affect the chick's future generations.

To get around this, Tizard and Doran looked to primordial germ cells (PGCs) — immature cells that eventually turn into sperm or eggs. Unlike in many animals, chicken PGCs spend time in the bloodstream during development. Researchers can therefore remove PGCs, edit them in the lab and then return them to the developing bird. The CSIRO team has even developed a method to insert CRISPR components directly into the bloodstream so that they can edit PGCs there⁴.

The researchers also plan to produce chickens with components required for CRISPR integrated directly into their genomes — what they call CRISPi chickens. This would make it even easier to edit chicken DNA, which could be a boon for 'farmaceuticals' — drugs created using domesticated animals.

Regulators have shown a willingness to consider such drugs. In 2006, the European Union approved a goat that produces an anti-clotting protein in its milk. It was subsequently approved by the US Food and Drug Administration, in 2009. And in 2015, both agencies approved a transgenic chicken whose eggs contain a drug for cholesterol diseases.





DE-EXTINCTION

About 4,000 years ago, hunting by humans helped to drive woolly mammoths (*Mammuthus primigenius*) to extinction. CRISPR pioneer George Church at Harvard Medical School in Boston, Massachusetts, has attracted attention for his ambitious plan to undo the damage by using CRISPR to transform endangered Indian elephants into woolly mammoths — or at least cold-resistant elephants. The goal, he says, would be to release them into a reserve in Siberia, where they would have space to roam.

The plan sounds wild — but efforts to make mammals more mammoth-like have been going on for a while. Last year, geneticist Vincent Lynch at the University of Chicago in Illinois showed that cells with the mammoth version of a gene for heat-sensing and hair growth could grow in low temperatures⁵, and mice with similar versions prefer the colder parts of a temperature-regulated cage⁶. Church says that he has edited about 14 such genes in elephant embryos.

But editing, birthing and then raising mammoth-like elephants is a huge undertaking. Church says that it would be unethical to implant gene-edited embryos into endangered elephants as part of an experiment. So his lab is looking into ways to build an artificial womb; so far, no such device has ever been shown to work.

There are some de-extinction projects that could prove less challenging. Ben Novak at the University of California, Santa Cruz, for example, wants to resurrect the passenger pigeon (*Ectopistes migratorius*), a once-ubiquitous bird that was driven to extinction in the late nineteenth century by overhunting. His group is currently comparing DNA from museum specimens to that of modern pigeons. Using PGC methods similar to Doran's, he plans to edit the modern-pigeon genomes so that the birds more closely resemble their extinct counterparts.

Novak says that the technology is not yet advanced enough to modify the hundreds of genes that differ between modern and historic pigeons. Still, he says that CRISPR has given him the best chance yet of realizing his lifelong dream of restoring an extinct species. "I think the project is 100% impossible without CRISPR," he says.

VECTOR CONTROL

For decades, researchers have explored the idea of genetically modifying mosquitos to prevent the spread of diseases such as dengue or malaria. CRISPR has given them a new way to try.

In November, molecular biologist Anthony James of the University of California, Irvine, revealed a line of mosquitoes with a synthetic system called a gene drive that passes a malaria-resistance gene on to the mosquitoes' offspring⁷. Gene drives ensure that almost all the insects' offspring inherit two copies of the edited gene, allowing it to spread rapidly through a population.

Another type of gene drive, published last December⁸, propagates a gene that sterilizes all female mosquitoes, which could wipe out a population. The outbreak of mosquito-borne Zika virus in Central and South America has increased interest in the technology, and several research labs have begun building gene drives that could eliminate the Zika-carrying species, *Aedes aegypti*.

Many scientists are worried about unintended and unknown ecological consequences of releasing such a mosquito. For this reason, Church and his colleagues have developed 'reverse gene drives' — systems that would propagate through the population to cancel out the original mutations^{9,10}.

But Jason Rasgon, who works on genetically modified insects at Pennsylvania State University in University Park, says that although ecology should always be a consideration, the extent and deadliness of some human diseases such as malaria may outweigh some costs.

Mosquitoes are some of the easiest insects to work with, he says, but researchers are looking at numerous other ways to use gene drives, including making ticks that are unable to transmit the bacteria that cause Lyme disease. Last year, researchers identified a set of genes that could be modified to prevent aquatic snails (*Biomphalaria glabrata*) from transmitting the parasitic disease schistosomiasis¹¹.

BETTER FOOD PRODUCTION

Last November, after a lengthy review, the US Food and Drug Administration approved the first transgenic animals for human consumption: fast-growing salmon made by AquaBounty Technologies of Maynard, Massachusetts. Some still fear that if the salmon escape, they could breed with wild fish and upset the ecological balance.

To address such concerns, fish geneticist Rex Dunham of Auburn University in Alabama has been using CRISPR to inactivate genes for three reproductive hormones — in this case, in catfish, the most intensively farmed fish in the United States. The changes should leave the fish sterile, so any fish that might escape from a farm, whether genetically modified or not, would stand little chance of polluting natural stocks. "If we're able to achieve 100% sterility, there is no way that they can make a genetic impact," Dunham says. Administering hormones would allow the fish to reproduce for breeding purposes. And Dunham says that similar methods could be used in other fish species.

CRISPR could also reduce the need for farmers to cull animals, an expensive and arguably inhumane practice. Biotechnologist Alison van Eenennaam at the University of California, Davis, is using the technique to ensure that beef cattle produce only male or male-like offspring, because females produce less meat and are often culled. She copies a Y-chromosome gene that is important for male sexual development onto the X chromosome in sperm. Offspring produced with the sperm would be either normal, XY males, or XX females with male traits such as more muscle.

In the egg industry, male chicks from elite egg-laying chicken breeds have no use, and farmers generally cull them within a day of hatching. Tizard and his colleagues are adding a gene for green fluorescent protein to the chickens' sex chromosomes so that male embryos will glow under ultraviolet light. Egg producers could remove the male eggs before they hatch and potentially use them for vaccine production.

There are other ways that CRISPR could make agriculture more humane. Packing cattle into trailers or other small spaces often causes injuries, especially when the animals have long horns. So cattle farmers generally burn, cut or remove them with chemicals — a process

that can be painful for the animal and dangerous for the handler. There are cattle varieties that do not have horns — a condition called ‘polled’ — but crossing these breeds with ‘elite’ meat or dairy breeds reduces the quality of the offspring.

Molecular geneticist Scott Fahrenkrug, founder of Recombinetics in Saint Paul, Minnesota, is using gene-editing techniques to transfer the gene that eliminates horns into elite breeds¹². The company has produced only two polled calves so far — both male — which are being raised at the University of California, Davis, until they are old enough to breed.

IMPROVING PETS

Last September, the genomics firm BGI wowed a conference in Shenzhen, China, with micropigs — animals that grow to only around 15 kilograms, about the size of a standard dachshund. BGI had originally intended to make the pigs for research, but has since decided to capitalize on creation of the animals by selling them as pets for US\$1,600. The plan is to eventually allow buyers to request customized coat patterns.

BGI is also using CRISPR to alter the size, colour and patterns of koi carp. Koi breeding is an ancient tradition in China, and Jian Wang, director of gene-editing platforms at BGI, says that even good breeders will usually produce only a few of the most beautifully coloured and proportioned, ‘champion quality’ fish out of millions of eggs. CRISPR, she says, will let them precisely control the fish’s patterns, and could also be used to make the fish more suitable for home aquariums rather than the large pools where they are usually kept. Wang says that the company will begin selling koi in 2017 or 2018 and plans to eventually add other types of pet fish to its repertoire.

Claire Wade, a geneticist at the University of Sydney in Australia, says that CRISPR could be used to enhance dogs. Her group has been cataloguing genetic differences between breeds and hopes to identify areas involved in behaviour and traits such as agility that could potentially be edited¹³. Sooam Biotech in Seoul, best-known for a service that will clone a deceased pet for \$100,000, is also interested in using CRISPR. Sooam researcher David Kim says that the company wants to enhance the capabilities of working dogs — guide dogs or herding dogs, for example.

Jeantine Lunshof, a bioethicist who works in Church’s lab at Harvard, says that engineering animals just to change their appearance, “just to

satisfy our idiosyncratic desires”, borders on frivolous and could harm animal well-being.

But she concedes that the practice is not much different from the inbreeding that humans have been performing for centuries to enhance traits in domestic animals and pets. And CRISPR might even help to eliminate some undesirable characteristics: many dog breeds are prone to hip problems, for example. “If you could use genome editing to reverse the very bad effects we have achieved by this selective inbreeding over decades, then that would be good.”

DISEASE MODELS

Ferrets have long been a useful model for influenza research because the virus replicates in their respiratory tracts and they sometimes sneeze when infected, allowing studies of virus transmission. But until the arrival of CRISPR, virologists lacked the tools to easily alter ferret genes. Xiaoqun Wang and his colleagues at the Chinese Academy of Sciences in Beijing have used CRISPR to tweak genes involved in ferret brain development¹⁴, and they are now using it to modify the animals’ susceptibility to the flu virus. He says that he will make the model available to infectious-disease researchers.

Behavioural researchers are particularly excited about the prospect of genetically manipulating marmosets and monkeys, which are more closely related to humans than are standard rodent models. The work is moving most quickly in China and Japan. In January, for instance, neuroscientist Zilong Qiu and his colleagues at the Chinese Academy of Sciences in Shanghai published a paper¹⁵ describing macaques with a CRISPR-induced mutation in *MECP2*, the gene associated with the neurodevelopmental disorder Rett syndrome. The animals showed symptoms of autism spectrum disorder, including repetitive behaviours and avoiding social contact.

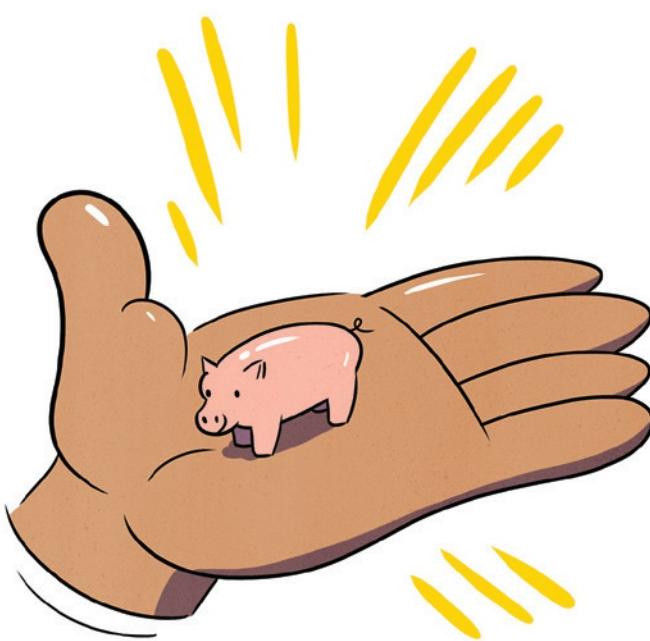
But Anthony Chan, a geneticist at Emory University in Atlanta, Georgia, cautions that researchers must think carefully about the ethics of creating such models and whether more-standard laboratory animals such as mice would suffice. “Not every disease needs a primate model,” he says.

Basic neuroscience could also benefit from the availability of new animal models. Neurobiologist Ed Boyden at the Massachusetts Institute of Technology is raising a colony of the world’s tiniest mammal — the Etruscan tree shrew (*Suncus etruscus*). The shrews’ brains are so small that the entire organ can be viewed under a microscope at once. Gene edits that cause neurons to flash when they fire, for instance, could allow researchers to study the animal’s entire brain in real time.

The CRISPR zoo is expanding fast — the question now is how to navigate the way forward. Pauwels says that the field could face the same kind of public backlash that bedevilled the previous generation of genetically modified plants and animals, and to avoid it, scientists need to communicate the advantages of their work. “If it’s here and can have some benefit,” she says, “let’s think of it as something we can digest and we can own.” ■

Sara Reardon writes for Nature from Washington DC.

1. Dhanapala, P., Doran, T., Tang, M. L. & Suphioglu, C. *Mol. Immunol.* **65**, 104–112 (2015).
2. Lillico, S. G. et al. *Sci. Rep.* **6**, 21645 (2016).
3. Whitworth, K. M. et al. *Nature Biotechnol.* **34**, 20–22 (2016).
4. Tyack, S. G. et al. *Transgenic Res.* **22**, 1257–1264 (2013).
5. Lynch, V. J. et al. *Cell Rep.* **12**, 217–228 (2015).
6. Marics, I., Malapert, P., Reynders, A., Gaillard, S. & Moqrich, A. *PLoS ONE* **9**, e99828 (2014).
7. Gantz, V. M. et al. *Proc. Natl Acad. Sci. USA* **112**, E6736–E6743 (2015).
8. Hammond, A. et al. *Nature Biotechnol.* **34**, 78–83 (2016).
9. DiCarlo, J. E., Chavez, A., Dietz, S. L., Esveld, K. M. & Church, G. M. *Nature Biotechnol.* **33**, 1250–1255 (2015).
10. Gantz, V. M. & Bier, E. *Science* **348**, 442–444 (2015).
11. Tennesen, J. A. et al. *PLoS Genet.* **11**, e1005067 (2015).
12. Tan, W. et al. *Proc. Natl Acad. Sci. USA* **110**, 16526–16531 (2013).
13. Arnott, E. R. et al. *Canine Genet. Epidemiol.* **2**, 6 (2015).
14. Kou, Z. et al. *Cell Res.* **25**, 1372–1375 (2015).
15. Liu, Z. et al. *Nature* **530**, 98–102 (2016).



COMMENT

GOVERNANCE Don't fear the DIY biologists, learn from them **p.167**

ECONOMICS China's fraught relationship with Latin America **p.169**

CONSERVATION Should half of Earth be set aside as wilderness? **p.170**

EPIDEMIOLOGY Standard strain-naming urgently needed for Zika **p.173**



PAULO FRIDMAN/BLOOMBERG VIA GETTY



In the United States, engineered crops now make up more than 80% of the soya bean (pictured), maize and cotton acreage.

Reboot the debate on genetic engineering

Arguments about whether process or product should be the focus of regulation are stalling progress, says **Jennifer Kuzma**.

Genetic engineering (GE) has become increasingly contentious in recent years. Thousands of citizens and stakeholders in the United States are currently striving to pass mandatory food-labelling laws, ban certain GE products and create GE-free zones for growing food.

GE is the manipulation of an organism's genome through biotechnology or modern molecular techniques. It is also called genetic modification, although that term is understood by scientists to encompass older processes such as hybridization as well. With the wealth of possibilities now offered by newly developed gene-editing tools — particularly CRISPR-Cas9 — debates about the safe and

appropriate uses of GE are becoming more heated. In fact, in the 20 years that I have been involved in discussions about it, oversight of GE has never seemed so much like a powder keg waiting to explode.

One issue that has dominated the debate is whether the focus of regulation should be the process by which GE organisms are made or the GE products themselves (the living organisms or products derived from them).

From 1999 to 2000, I directed a US National Academy of Sciences study (see go.nature.

com/lhyten) to investigate pest-resistant GE plants and their regulation. While working on that project, and in the years since, I have found that most people in favour of product-based regulation believe that there is no need to treat GE organisms differently from conventionally bred ones. Moreover, these people often claim that those who think that the process of engineering should be the focus of regulation — and thus, who want to see most or all GE products go through regulatory review before they enter the marketplace — are making arguments based on values or emotions, rather than science, to support their views.

But framing the debate around 'product versus process' is neither logical nor scientific. It is stalling productive ▶



CRISPR EVERYWHERE

A *Nature* special issue
nature.com/crispr

► dialogue on the development of appropriate oversight in the face of rapid advances in GE.

IN A RUT

The United States has had a system in place for overseeing GE products since the mid-1980s: the Coordinated Framework for Regulation of Biotechnology (CFRB). The parties involved in the development of this framework — including representatives from the Office of Science and Technology Policy (OSTP) and various federal agencies — determined that it is the final product of GE that potentially poses a risk to human health and the environment, not the process by which the product is made¹.

Product-led regulation was seen to be a science-based approach that would preclude the need for new biotechnology laws. It meant that GE organisms could be covered by existing laws for products intended to be used as pesticides, plant pests, toxic substances and so on; engineered organisms could be channelled to particular agencies — the Environmental Protection Agency (EPA), the Food and Drug Administration (FDA) and the US Department of Agriculture (USDA) — depending on what category they fell into.

So the intended use of a product has dictated which agency has the authority to regulate it under the CFRB. Yet, in practice, it is the process of GE that has been the ‘regulatory trigger’ used to capture products for pre-market review.

After the CFRB was published in 1986, each agency produced documents that detailed the specific protocol for the GE-product categories under its purview. For example, the EPA described the steps that developers would need to take if they were marketing plants that have pesticide-like substances engineered into them, whereas the USDA laid out how developers should handle GE plants considered to be ‘plant pests’.

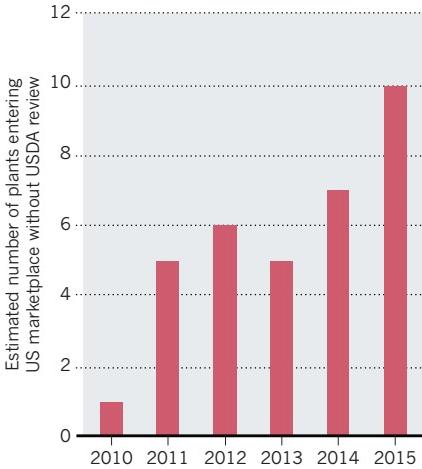
These EPA and the USDA documents specified that organisms made by recombinant-DNA technologies or GE (but not their conventionally bred counterparts) must go through regulatory review before entering the marketplace.

The FDA took a different approach. It recommended through a guidance document — not a regulation — that developers of foods derived from ‘new plant varieties’ undergo a voluntary consultation process with the agency. This guidance did not exclude non-GE new plant varieties. In practice, however, developers of conventionally bred foods seem not to have undergone such consultations, whereas the FDA has been notified of more than 100 foods derived from GE plants (see go.nature.com/z78s1e).

For the EPA, the USDA and the FDA, the

LOOSER SCRUTINY

Because of changes to genetic-engineering (GE) processes, several GE crops have entered the US marketplace without review from the US Department of Agriculture (USDA) in recent years.



engineered product once again becomes the focus when the agencies actually assess the level of risk that it poses. But from a scientific standpoint, a product’s traits — harmful or otherwise — depend in part on the process by which it is made. (This is especially evident from human gene-therapy trials, where new methods for delivering genes have removed the need for potentially harmful viral vectors.) And in their review procedures, the agencies recognize that the process of engineering is important. The USDA, for example, requires a “detailed description of the molecular biology of the system ... used to produce the regulated article”.

Thus, product and process issues are not distinct in regulation. Indeed, it does not make sense scientifically to try to value one approach more highly than the other.

The idea that regulating products is the only ‘science-based’ way has been popular with regulators and developers beyond the United States. For instance, plant scientist Ingo Potrykus, who led the development of the genetically engineered vitamin-A-enriched ‘golden rice’ variety at the Swiss Federal Institute of Technology (ETH) in Zurich, stated in 2010 that it would be a “crime against humanity” not to change from “regulating a technology on ideological terms” to “science-based regulation, guided by considerations of the risks and benefits of the trait”².

Yet many countries go further than the United States when it comes to process-based triggers for regulation, relying on national laws. In Brazil, a national biosafety law provides safety standards and oversight mechanisms for GE organisms; in Australia, the Gene Technology Act mandates a regulatory framework for the risk assessment and management of GE organisms.

EMERGING TECHNOLOGIES

The product-versus-process framing has reared its confusing head again in recent discussions. Gene editing involves changing DNA sequences at targeted locations, usually using site-directed nucleases (proteins that naturally cut DNA), such as CRISPR-Cas9, TALENS and zinc-finger nucleases. With these tools, genetic engineers can introduce one or a few nucleotide changes to a gene, make insertions or deletions in a gene sequence, or insert a different gene altogether, potentially from a different species. International discussions have focused on which types of gene-editing manipulation fall under regulatory definitions of GE organisms in different oversight regimes³.

Ironically, the same GE developers who once claimed that the process of GE does not matter for regulatory purposes are now arguing that changes to the engineering process justify looser regulatory scrutiny⁴. They contend that gene editing is a safer process than first-generation GE techniques owing to its precision and the smaller point mutations often made⁵.

And some US regulatory agencies are heeding these calls. Thanks to emerging methods of gene delivery and gene editing, genetic engineers no longer need to use DNA sequences from plant pests to introduce engineered genes into host plants. In part because of this change to the process by which the organisms are being made, the USDA has, for about five years, decided not to regulate about 20 engineered plants (see ‘Looser scrutiny’). Several have entered the market without going through any formal regulatory review — either by the USDA or other agencies.

In Europe, crop developers are anxiously waiting for the European Commission to decide how changes to GE processes should

“It is impossible to be completely ‘science based’ in a regulatory system.”

affect regulatory policy. Specifically, the commission is expected soon to deliver a verdict on whether the definition of GE organisms covers gene-edited plants in which any foreign DNA used in the engineering process has been removed through selective breeding — and which are indistinguishable from wild plants that might have acquired the same mutation naturally (see *Nature* 528, 319–320; 2015).

GE developers and some regulators have been inconsistent in their product-versus-process arguments for good reason. The dichotomy doesn’t work, in practice or in theory. In fact, product-based arguments lead to one of two conclusions: if all products (GE or otherwise) are to be treated the same, then either all products — GE and conventionally bred — should be regulated,

or neither should be. The first option is impractical and the second inadvisable given that some products could be harmful.

A FRESH START

It is time to reset the debate. Product-versus-process arguments reflect world views about the desired level of regulation for GE organisms. These underlying viewpoints should be made explicit, and the idea that product-based regulation is the only science-based approach rejected.

In reality, it is impossible to be completely 'science based' in a regulatory system. Value judgements are embedded in all risk and safety assessments. For example, the dose-response curve for a certain food additive might be known, but such data do not by themselves tell regulators where to set an acceptable safety limit. More often, the dose-response curve is not well established, or known at all. This uncertainty leads to various interpretations of the data.

Empirical evidence matters, but human interpretation brings meaning to that evidence, and multiple perspectives can strengthen understanding. Thus, an oversight system should focus on what concerns a diversity of stakeholders and citizens have,

what evidence or risk-mitigation strategies can help to address those concerns, and what classes of GE products or processes should receive greater regulatory scrutiny. In practice, regulators and other stakeholders will need to consider a mix of product and process issues to capture product groups that are likely to be of greater concern.

Several models in the social-science literature describe how such democratic deliberation might be achieved⁶. And Norway's decision-making about GE organisms under its gene-technology act demonstrates how factors outside 'science-based' health or environmental harms can be incorporated into formal regulatory processes in practice. Since 2005, regulators in Norway making decisions about whether a GE organism will be released into the environment consider the results of safety reviews, and whether participants of a consultation process perceive that the organism provides a better option than alternatives and contributes to sustainable agricultural practices (see go.nature.com/5nxzcn).

There is a chance to start over, in the United States and elsewhere. In part because of advances in gene editing and a greater diversity of GE organisms being presented to regulators, the OSTP initiated a process in

July 2015 to clarify which regulatory authority is responsible for what under the CFRB⁷. And just last month, the USDA published four possible scenarios for a proposed new framework for the regulation of GE crops⁸.

Within these efforts and others, stakeholders could do away with polarizing product-versus-process and science-versus-values framings, and help to establish a governance system that is both informed by the science and guided by the concerns and values of citizens. ■

Jennifer Kuzma is distinguished professor in the social sciences and co-director of the Genetic Engineering and Society Center at North Carolina State University, USA.
e-mail: jkuzma@ncsu.edu

1. OSTP. *Fed. Reg.* **51**, 23302 (1986).
2. Potrykus, I. *N. Biotechnol.* **27**, 466–472 (2010).
3. Wolt, J. D., Wang, K. & Yang, B. *Plant Biotechnol. J.* **14**, 510–518 (2015).
4. Huang, S., Weigel, D., Beachy, R. N. & Li, J. *Nature Genet.* **48**, 109–111 (2016).
5. Kokotovich, A. & Kuzma, J. *Bull. Sci. Technol. Soc.* **34**, 108–120 (2014).
6. Ramachandran, G. et al. *J. Nanopart. Res.* **13**, 1345–1371 (2011).
7. Waltz, E. *Nature Biotechnol.* **33**, 1221–1222 (2015).
8. USDA. *Fed. Reg.* **81**, 6225–6229 (2016).

Learn from DIY biologists

The citizen-science community has a responsible, proactive attitude that is well suited to gene-editing, argues Todd Kuiken.

One of the top science stories of 2012 involved a furore about the wisdom of enhancing the transmissibility of the H5N1 avian influenza virus in ferrets. In that same year, fears mounted that do-it-yourself (DIY) biologists would cook up their own versions of the virus using information published in the academic press.

Now, journalists and others are again targeting the citizen-science community — a group of people with or without formal training who pursue research either as a hobby or to foster societal learning and open science — amid fears about the nascent gene-editing technology CRISPR–Cas9. In January, the *San Jose Mercury News* ran an article under a pearl-clutching headline: "Bay Area biologist's gene-editing kit lets do-it-yourselfers play God at the kitchen table." And although they are much less alarmist, scholars are advising policymakers to consider the potential uses of gene editing "outside the traditional laboratory setting" (R. A. Charo & H. T. Greely *Am. J. Bioeth.* **15**, 11–17; 2015).

The reality is that the techniques and

expertise needed to create a deadly insect or virus are far beyond the capabilities of the typical DIY biologist or community lab. Moreover, pursuing such a creation would go against the culture of responsibility that DIY biologists have developed over the past five years. In fact, when it comes to thinking proactively about the safety issues thrown up by biotechnology, the global DIY-biology community is arguably ahead of the scientific establishment.

EASY ACCESS

The equipment and reagents that are needed to use CRISPR–Cas9 are already readily available to DIY biologists. Members of the teams that participated in the 2015 International Genetically Engineered Machine (iGEM) competition — including high-school students and users of community labs around

the world — received CRISPR–Cas9 plasmids in their starting kits. These kits contain more than 1,000 standard biological parts known as BioBricks, the DNA-based building blocks that participants need to engineer a biological system for entering into the competition. Other components of the CRISPR–Cas9 system are also available from the iGEM registry (<http://parts.igem.org/CRISPR>).

Yet few DIY biologists seem to be using the technology. Both Tom Burkett, founder of the Baltimore Under Ground Science Space in Maryland, and Ellen Jorgensen, executive director of Genspace — a community lab in Brooklyn, New York — say that their users are interested in CRISPR–Cas9, and Genspace will be offering a workshop on it in March. But none of the projects currently being pursued in these spaces require it. Users of the La Paillasse community lab in Paris are similarly focused on projects that do not need CRISPR–Cas9.

The materials might be available, but the knowledge and understanding needed to make edits that have the desired effects





Users of the Baltimore Under Ground Science Space are not yet using CRISPR-Cas9.

are not. Also, most DIY biologists are interested in building genetic circuits in bacteria or yeast, and they can generally do this using well-established techniques, such as SLiCE (seamless ligation cloning extract), and with genes that have been synthesized by commercial suppliers or that can be obtained from the iGEM registry.

CRISPR-Cas9 is a fast-moving technology that may well become more popular with DIY biologists in the coming months and years. Even if this happens, there is no *a priori* reason to expect this community to cause more harm when using it than anyone else.

GOOD CONDUCT

The DIY-biology community developed codes of conduct in mid-2011 (<https://diybio.org/codes>). At this point, the community comprised one shared laboratory (Genspace), which opened in December 2010, and a loose-knit collection of groups from across the globe, each with different levels of expertise, resources and protocols.

In discussions online and in face-to-face gatherings, it emerged that if the DIY-biology community was to advance and start pursuing more-sophisticated projects, it would need to develop a set of governance principles. I and Jason Bobe, a co-founder of DIYbio.org, an online hub for people interested in pursuing DIY biology, convened a series of workshops that brought together groups from the United Kingdom, Denmark, France and Germany. We then repeated the exercise with six groups in the

United States. We knew that a set of rules that outlined appropriate practices would be effective only if such rules had been developed and agreed on together.

Today, Genspace and other community labs around the world have their own advisory boards or can seek advice from the 'Ask a biosafety professional your question' portal (<http://ask.diybio.org>). The portal's panels review proposals for projects and flag potential safety issues. In the United States, community labs have even developed relationships with the Federal Bureau of Investigation, which has introduced members to local police and fire departments to maximize preparedness for security issues that could arise.

In many ways, this proactive culture of responsibility is an advance on the post hoc scrambling that often occurs within the scientific establishment. Much of the debate about the pros and cons of the H5N1 experiments took place while the work was under review for publication.

And in the case of gene editing, even the US National Academy of Sciences was caught on the hop. It did not begin to seriously discuss the risks associated with using the approach to engineer genes that could quickly spread through wild populations — known as gene drives — until after experiments demonstrating the concept in fruit flies had been published in a peer-reviewed journal (V. M. Gantz & E. Bier *Science* **348**, 442–444; 2015).

Of course, community norms will have little effect on the behaviour of rogue

individuals who are intent on causing mischief or harm. But such people could just as easily be scientists working in government, university or commercial labs as DIY biologists. Indeed, the current culture of responsibility among DIY biologists, their collaborative style of working and the fact that community labs are open spaces in which everyone can see what is going on reduce, if not eliminate, doomsday scenarios of mutant organisms escaping from basements and causing harm.

One development that has increased anxiety about the use of CRISPR-Cas9 by DIY biologists is a crowdsourcing venture by synthetic biologist Josiah Zayner, founder of the Open Discovery Institute in Burlingame, California. Thirty days after launching his campaign on the crowdfunding website Indiegogo last November, Zayner had raised almost US\$34,000 to fund the production and distribution of DIY CRISPR kits — supposedly to help people "learn modern science by doing". (He has since raised more than \$62,000, six times his original goal.)

But the concern about Zayner's project arises not because it gives people outside conventional labs more capabilities than they would otherwise have had. DIY biologists already use various tools to assemble DNA fragments in bacteria and yeast — the microorganisms that he supplies in his kits. Zayner's campaign is worrisome because it does not seem to comply with the DIYbio.org code of conduct. The video that accompanies his campaign zooms in on Petri dishes containing samples that are stored next to food in a refrigerator. More than anything, Zayner's campaign is a reminder of the myriad ways in which researchers — conventional or otherwise — can now get their work funded.

With the ready availability of tools such as CRISPR-Cas9 and crowdfunding, a more-decentralized governance is needed for everyone, not just DIY biologists. Codes of conduct will be needed to establish appropriate norms for government funding and regulatory agencies, for people working both within and outside conventional research settings, for the directors of community labs and for the developers of crowdfunding platforms.

The DIY-biology community, as a stakeholder that has already addressed many of the underlying issues, should take part in a robust public dialogue about the use of CRISPR-Cas9 and how governance models can ensure safe, responsible research. ■

Todd Kuiken is a senior program associate and principal investigator of the Wilson Center's Synthetic Biology Project in Washington DC.
e-mail: todd.kuiken@wilsoncenter.org



Protests against China's Las Bambas mining project in Peru erupted into violence last year.

ECONOMICS

China in the new world

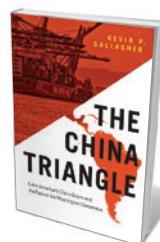
Margaret Myers on a study of the impacts of the country's presence in Latin America.

For nearly 15 years, China's economic and construction boom, coupled with high commodity prices, has driven soaring demand for steel, copper and other commodities. Latin America has been a crucial partner, with Chile, for instance, accounting for 40% of raw copper imports and Brazil responsible for 49% of soya imports. China's mounting concerns about food and energy security have also prompted engagement with resource-rich Venezuela, Peru and Argentina. The China-fuelled 'super-cycle' has left a path of environmental destruction and social conflict in Latin America. Now, China's economic slowdown and a slump in commodity prices spell bleaker economic prospects for exporting nations. For instance, Chilean copper company Codelco announced massive layoffs in 2015 as copper prices dropped.

As one of the first accounts of post-'China boom' Latin America, Kevin Gallagher's *The China Triangle* adds much to a profusion of books on China–Latin America relations. By skilfully framing Latin America's development challenges — such as lack of highly skilled labour — in a historical context, Gallagher reminds us that commodity-led growth is hardly a new phenomenon in the

region. The end of the nineteenth century saw the first boom, when Europe and the United States began to import raw materials in serious quantities. Years of dependence on exporting natural resources led to wide-ranging policy outcomes in Latin America, from state-led industrialization to the Washington Consensus, a set of US prescriptions for economic development in the region in the 1990s and 2000s. *The China Triangle* is largely premised on the idea that the most recent phase in Latin America's economic development was as much a rejection of the consensus as an embrace of China.

As Gallagher shows, that relationship has been rocky in many ways. Gallagher documents the negative effects of booming trade and investment by Chinese and other firms in extraction of natural resources. Growth



The China Triangle: Latin America's China Boom and the Fate of the Washington Consensus
KEVIN P. GALLAGHER
Oxford University Press: 2016.

in mining alone has led to increased chemical leaching, improper disposal of waste and acidic runoff from mines. Chemical leaching has killed fish and caused economic damage in Peru. Deforestation and related flooding in Argentina are thought to stem from the rise in soya production for export to China.

Gallagher is careful to note that Chinese companies have shown their capacity to adapt to Latin American laws and norms. In Peru, US company Doe Run performed much worse on a number of counts than Chinese mining firms. But the vast majority of trade, 90% of Chinese investment in Latin America and 80% of its loans to the region's governments are focused in sectors linked to environmental degradation. Hence China has, on average, more environmental impact in Latin America than do other partners.

Concerns surround China's hydroelectric-dam projects in the region, including the Coca-Codo Sinclair dam, Ecuador's largest energy project. Although this is expected to address a critical energy deficit, many Ecuadoreans are concerned about water diversion from the San Rafael Falls, a prominent tourist destination, and the construction of access roads in the Amazon. The Néstor Kirchner ▶

► and Jorge Cepernic dams in Argentina were touted as key energy projects by former president Cristina Fernández de Kirchner, but they are far from the grid and about 2,750 kilometres from Buenos Aires, where energy needs are high. A 2006 feasibility study of 30 dam projects by Argentina's energy ministry ranked them 23rd and 25th, respectively.

In 2015, Chinese company MMG Ltd modified its environmental-impact study for the Las Bambas copper-mining project in Peru's Cotabambas province without consulting local communities. Although compliant with newly modified Peruvian law, the decision provoked demonstrations by local residents that ended in 4 deaths and led Peruvian President Ollanta Humala to declare a 30-day state of emergency in the province.

Despite China's slowing growth and some bad press, the country will — Gallagher reveals — remain one of the region's key economic partners. Last year saw Chinese finance to Latin America and the Caribbean rise to a level surpassed only in 2010, much of it focused on oil, gas and transport infrastructure. Just as US investors did in the decades straddling the turn of the last century, China is seeking to develop transport networks to carry commodities to port, such as the Peru–Brazil railway proposed during Chinese Premier Li Keqiang's 2015 visit to the region.

Latin America also stands to benefit from China's sustained presence. In *The Dragon in the Room* (Stanford University Press, 2010), Gallagher and Roberto Porzencanski advised nations to capture China's windfall by investing in export diversification. They did not, but Gallagher insists in *The China Triangle* that it is not too late. He prescribes greater partnerships between countries and markets, and policies that promote equality and environmental stewardship. But post-boom, Gallagher foresees a Latin America with less room to manoeuvre, economically and politically. The region would need to appeal to both the United States and China to ensure future growth. Gallagher's 'China triangle' refers to this shift.

The value of diversified partnerships, whether with the United States and China or a wider variety of partners, is increasingly clear to Latin Americans. The region should avoid dependency on raw-materials exports — and beware of reliance on big powers with deep pockets. ■

Margaret Myers is director of the China and Latin America programme at the Inter-American Dialogue, a Western Hemisphere affairs think tank in Washington DC.
e-mail: mmyers@thedialogue.org

CONSERVATION

Glass half full

Stuart Pimm examines E. O. Wilson's grand vision for an Earth shared equally between humanity and nature.



Biologist E. O. Wilson suggests a radical approach to conservation.

What do humans want? So asks E. O. Wilson near the start of *Half-Earth*, his bold vision for the biosphere. He outlines the probable answer: "indefinitely long and healthy life for all, abundant sustainable resources, personal freedom, adventure both virtual and real on demand, status, dignity, membership in one or more respectable groups, obedience to wise rulers and laws, and lots of sex with or without reproduction". He adds: "These are also the goals of your family dog."

The eminent biologist demands that we

aspire to much more. He calls for no less than committing half of the planet's surface to a haven for nature. It's an ambitious goal, yet failure would be dire.

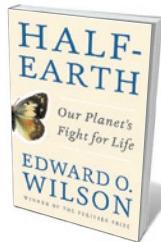
This is no isolated argument. *Half-Earth* is the last in a trilogy. In *The Social Conquest of Earth* (Liveright, 2012), he marvels at how advanced social organization is rare among animals and how "one species of large-sized African primate" — us — has become not merely dominant, but a force of geological change. In *The Meaning of Human Existence* (Liveright, 2014), he argues that we are

"a biological species in a biological world", adapted to going forth and multiplying as if there were no tomorrow. There might not be. Only if we protect 'Half-Earth' can the vast majority of species be saved.

Wilson's vision begs questions that he does not address in detail. Is it feasible? How close are we to achieving it? In which ecosystems — forests or deserts or reefs — might we succeed? Where might failure be inevitable? Instead, he presents a manifesto. Half, he says, is a safe limit, because our own survival depends on the services of nature. Wilson argues a psychological need, too. He intends his goal to inspire us to strive nobly against the odds on behalf of all life. We must articulate an endpoint beyond the day-to-day business of saving particular species and habitats.

The consequences of protecting less than half are as close as my local supermarket on Key Largo, Florida, where I do my fieldwork. It is 500 metres from the Atlantic Ocean on one side and the Gulf of Mexico on the other, yet the fishmonger's slab is covered with farm-raised salmon and tilapia, and scallops from the Southern Hemisphere. Even the mahimahi — available locally — is from Mexico.

Wilson castigates those who think that there is no problem with humans eliminating species 1,000 times faster than the natural background rate. Will new species evolve as they did after the mass extinction that killed the non-avian dinosaurs? It took evolution 5 million years to restore previous



Half-Earth: Our Planet's Fight for Life
EDWARD O. WILSON
Liveright: 2016.

levels of diversity. Will invasive species fill in the gaps? Alien species from rabbits in Australia to zebra mussels in the United States already cause harm costing billions of dollars per year.

Nor is Wilson kind to "new conservation", a movement that he notes is embraced by the large US land trust the Nature Conservancy. Its proponents denigrate those who believe in pristine landscapes and, as he puts it, "prefer 'working landscapes' presumably as opposed to 'lazy and idle' landscapes, thereby making them more acceptable to ... business leaders". A Google search suggests that the term pristine landscapes may have appeared in the flagship journal *Conservation Biology* once in the past decade — raising the question of who the professionals are who supposedly believe in them.

The Amazon exemplifies what Wilson calls wilderness: regions with small human populations, mainly indigenous ones. Companies that extract resources have historically been insensitive to the cultural disruption, and even genocide, that this can trigger. Wilson emphasizes how cultural diversity and biodiversity are important and can reinforce each other. I share his impression that the individuals most uncaring and dismissive of wilderness and biodiversity are those who have had the least experience of it. As nineteenth-century explorer Alexander von Humboldt put it: "The most dangerous worldview is the worldview of those who have not viewed the world."

Is Half-Earth possible? The trajectories are favourable. About 5 million square kilometres of land and almost none of the oceans were protected in the mid-1970s; now the figures are close to 17 million and 10 million square kilometres, respectively. Vast marine no-take zones have been established

annually since 2000. Globally agreed targets aspire to more, and more representative, protection. Large tracts of land — deserts, the Amazon, the boreal forests — are protected because they are remote. The challenge will be to protect areas near cities, or areas that, like temperate grasslands, are easy to convert to livestock grazing.

A change in moral reasoning gives Wilson most hope. A 2015 encyclical letter from Pope Francis contains an outstanding tour of the challenges in mitigating damage to natural habitats. Its moral imperative, that we have no right to do harm, echoes Wilson's concluding sentence: "Do no further harm to the biosphere."

Wilson lauds those who devote their lives to that cause. The degraded longleaf-pine savannahs of the US Gulf coast — neglected by

"We must articulate an endpoint beyond saving particular species and habitats."

federal authorities and land trusts — found a champion in the philanthropist M. C. Davis. Entrepreneur Greg Carr has helped to restore Gorongosa National Park in

Mozambique after a brutal civil war. Entrepreneurs Douglas and Kristine Tompkins have protected more land worldwide than any other private individuals — and in temperate grasslands, to boot. Progress on Half-Earth is possible in unlikely places. It is an aspiration worthy of our species. ■

Stuart Pimm is professor of conservation at the Nicholas School of the Environment at Duke University in Durham, North Carolina, and directs the non-profit organization SavingSpecies.
e-mail: stuartpimm@me.com

PSYCHOLOGY

No blank slate

Sara Reardon is moved by a play about the toll of infant sex-assignment surgery.

In 1966, psychologist John Money of Johns Hopkins University in Baltimore, Maryland, met someone who he felt was the research patient of a lifetime. David Reimer, then an eight-month-old boy, had had his penis mutilated in a circumcision accident. Doctors concluded that surgical reconstruction was too difficult. Money proposed a 'solution': could the child be turned into a girl?

Money studied people born with intermediate sex characteristics — then called

hermaphrodites. Standard medical procedure at that time (and still all too often) was to guess the sex that a baby 'should' be and surgically alter their genitals accordingly.

Money believed, as did many psychologists at the time, that the right training and environment could shape a child into any gender as long the process was started early enough. And because there was no doubt about whether

Boy
ANNA ZIEGLER
Clurman Theatre, New York City.
Until 9 April 2016.

Reimer had been born a boy, and without the other variables such as hormonal or genetic characteristics that can contribute to gender identity in intersex individuals, Money thought that the baby presented the perfect test case for the nurture theory. Reimer even had a control, a twin brother.

This tragic experiment is the inspiration for Anna Ziegler's play *Boy*, now showing at the Clurman Theatre in New York City. The story cuts between 1989 and the 1970s, following the young adulthood of Adam



Bobby Steggert (left) as Adam and Paul Niebanck as psychologist Wendell Barnes in *Boy*.

(played by Bobby Steggert) and his childhood, first as a baby boy named Samuel, then as a girl. After a circumcision accident, Samuel's parents reach out to famous psychologist Wendell Barnes (Paul Niebanck), who counsels them that being an "incomplete" male would do irreparable damage to the child's psyche, whereas raising him as a girl should be fine. "We're blank slates," says Barnes. Thus, Samuel becomes Samantha.

After the child is given surgery to create a vagina, Barnes imposes a harsh regimen of counselling and hormone treatments. Samantha must never know the circumstances of her birth: Barnes believes that the revelation would scar her. He regularly flies the family from Iowa to meet with him in Massachusetts. Samantha's mother is given a script and directed to overload her child with stereotypical female interests: dolls, baking, figure skating, "open conversations about our bodies".

Steggert hops back and forth across the stage, alternating between the adult Adam and the child Samantha with neither a costume change nor a major personality shift. His constant face is a powerful reminder that the same male person is there all

NATURE.COM
For more on science
in culture see:
[nature.com/
booksandarts](http://nature.com/booksandarts)

along, even if Samantha temporarily accepts girlhood.

What we don't see immediately is how broken she is. Being kissed by a boy in junior high school was the worst experience of his life, the adult Adam tells his girlfriend Jenny. Samantha would urinate standing up, try to shave her face and watch her pubescent brother explore his own sexuality in their shared bedroom. "I had it too, this sensation of wanting to get somewhere," the adult Adam explains in a heartbreak letter to Barnes. "But I'd look down and there was nothing there."

The child Samantha tells Barnes none of this; she is desperate to please the doctor who she believes cares more for her than her own parents. Barnes listens when her parents do not and plays chess with her while every child in her class shuns her. She begs to move in with him.

Their special relationship breaks down when Samantha enters puberty. Barnes insists that she undergo surgery to repair the "vagina you were born with". He urges: "You need to be made whole."

Today, intersex advocates decry such language, arguing that people can live happy lives with uncommon genitalia. Whether to have surgery, they feel, should be a choice made by the person themselves, as an adult.

So legal challenges are beginning to mount. In 2013, a couple sued the Medical University of South Carolina in Charleston for performing surgery on their adopted son, who had been wrongly assigned female genitalia before they adopted him. And in 2015, the United Nations' special rapporteur on torture called newborn genital surgery a human rights violation — a follow-up report will be released this year.

But change is a slow process. And Money's work was groundbreaking, despite all the harm that it caused to Reimer — a failure that devastated the psychologist. He was among the first to describe sex — defined by physical traits — as a distinct entity from gender, which is how one identifies oneself. Feminists seized on his work as proof that women's difficulties in typically male professions are the result of culture rather than biology. And Money, who died in 2006, supported surgery for older people who felt that they had been assigned the wrong sex.

So the writers of *Boy* deserve credit for not portraying a stereotypical arrogant scientist willing to do anything to prove his theory, an accusation that Samantha's parents make. Although clearly eager to defend his work — who isn't? — Barnes does seem to care for Samantha. He writes to her frequently, teaches her classic literature and becomes genuinely distressed by her problems at home and school. "Not only is [she] an exemplar for science, she is a delightful girl," he tells the audience at a lecture. In their final showdown, when Adam confronts Barnes and reveals that he had his penis reconstructed at age 15, Barnes accepts Adam's decision and admits that he is male.

Despite some tedious dialogue, we cringe at the physicality of Adam's struggle each time he considers whether to share his secret with his love, even as every instinct screams no. Afraid to touch Jenny, Adam focuses his attention on her toddler son — the child he desperately wants but will never be able to have, no matter how much reconstructive surgery he undergoes.

Adam's love story ends predictably, but his future is probably far from rosy. Money was not wrong about the incredible malleability of children. Although they are far from being blank slates, children are perhaps like line drawings, coloured in by experience. Adam's 15 years of lies, sexual confusion, hormone treatments and social exclusion will not be easily overcome.

David Reimer found love, marrying in 1990. But 14 years later, at the age of 38, he killed himself. He was a victim of a rush to put children into neatly labelled buckets that continues even today. ■

Sara Reardon is a staff reporter for Nature in Washington DC, writing about biomedical research and policy.

Correspondence

Zika virus: designate standardized names

A rapid response by the public-health and research communities to infectious viral diseases depends on the reproducible tracking and analysis of pathogen isolates. A standard strain-naming convention for Zika virus sequences is therefore urgently needed. This will ensure that the exchange and interpretation of data is unambiguous in efforts to contain the current outbreak in the tropical Western Hemisphere.

Zika virus strain names for isolates associated with the outbreak are arbitrarily designated as BeH818995, ZikaSPH2015 and BR/949/15, for example. Such names are largely opaque and inconsistent when it comes to context, although some may include useful metadata about isolates. It is impractical to include all relevant metadata in the isolate name, but some consistent information is useful for identifying specific isolates.

Building on conventions in other viral fields, we urge the Zika community to adopt a standard nomenclature for isolate names, specifying the virus type (ZIKV), host species abbreviation, geographical location of isolation, unique identification string and year of isolation. The preferred isolate name for BeH818995, for example, would then be ZIKV/H. sapiens/Brazil/BeH818995/2015.

Richard H. Scheuermann*
J. Craig Venter Institute, La Jolla, California, USA.

rscheuermann@jcv.org
*On behalf of the Viral Genome Annotation Standards Working Group (see go.nature.com/i5dewk for full list).

Zika virus: accurate terminology matters

You describe microcephaly as a “serious congenital malformation”, which risks confusing the public and causing

needless distress to the families of children with small heads, irrespective of whether these are linked to Zika virus infection (see *Nature* **530**, 5; 2016). In fact, ‘microcephaly’ simply means a small head and is not necessarily associated with intellectual disability, as is often assumed.

Microcephaly is a feature of hundreds of different conditions, but can also be seen in otherwise normal individuals (P. Merlob *et al.* *J. Med. Genet.* **25**, 750–753; 1988; S. Ashwal *et al.* *Neurology* **73**, 887–897; 2009).

This is not mere semantics. Investigations into the proposed link between Zika virus and birth defects (for which there seems to be little evidence at present) will need to include systematic assessment of all the possible causes of microcephaly in children thought to have been affected by the virus (C. G. Victora *et al.* *Lancet* **387**, 621–624; 2016).

Edwin P. Kirk *Sydney Children's Hospital; University of New South Wales; and SEALS Laboratories, Randwick, Australia.*
edwin.kirk@health.nsw.gov.au

How to engage social scientists in IPBES

We contend that the disciplinary imbalance within the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) could best be remedied by improving the organization's communication with researchers from the social sciences and humanities (see A. B. M. Vadrot *et al.* *Nature* **530**, 160; 2016).

Our analysis of the groups that were nominated and selected after the second IPBES call for experts for deliverables 2(b) and 3(b)(i) — namely the regional/subregional assessments of biodiversity and ecosystem services, and of land degradation and restoration — indicated that most people who applied for the assessments had a background in natural

sciences (see go.nature.com/pexril). This suggests that IPBES communications about the details and implications of the IPBES process itself might not be effectively engaging the social-science and humanities communities.

We suggest that IPBES calls need to be circulated more widely and avoid language and expressions that are tailored specifically for natural scientists. The calls should recognize differences in the social-science and humanities communities and target these more specifically.

Katrin Reuter, Malte Timpte
Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany.

Carsten Neshöver *Helmholtz Centre for Environmental Research — UFZ, Leipzig, Germany.*
malte.timpte@mfn-berlin.de

Better management of alien species

In our view, the European Union's recent legislation on invasive alien species will be an effective conservation tool only if the inclusion of new species is supported by the majority of EU states. We call for Europe to put the protection of its biodiversity before the short-term economic interests of member states.

Europe is one of the world's most biologically invaded regions (M. van Kleunen *et al.* *Nature* **525**, 100–103; 2015). But the list of invasive alien species targeted for action under the January 2015 EU legislation includes just 37 entries (see go.nature.com/gigftz) — even though Europe hosts more than 1,000 such species, most of which meet the criteria for listing (M. Vila *et al.* *Front. Ecol. Environ.* **8**, 135–144; 2010). For example, knotweed (*Fallopia* sp.) and American mink (*Neovison vison*) are well-characterized species that are responsible for extensive biodiversity losses across the continent.

We are concerned that the restricted new listing cannot hope to address the scale of biological invasions in Europe. Management must be coordinated at the EU level if both protective and preventative regulation are to be widely applicable, comprehensive and effective.

Jan Pergl *Institute of Botany, The Czech Academy of Sciences, Pruhonice, Czech Republic.*

Piero Genovesi *Institute for Environmental Protection and Research, Rome, Italy.*

Petr Pyšek *Institute of Botany, The Czech Academy of Sciences, Pruhonice; and Charles University in Prague, Czech Republic.*
jan.pergl@ibot.cas.cz

Class uncorrected errors as misconduct

Post-publication peer review is becoming increasingly popular, but authors need more incentive to self-correct and amend the scientific record (see D. B. Allison *et al.* *Nature* **530**, 27–29; 2016). We propose that failure by authors to correct their mistakes should be classified as scientific misconduct. This policy has already been implemented by our institute, and we encourage research institutions and funding bodies to follow suit (see go.nature.com/dgiff).

The responsibility to correct errors lies mainly with the criticized authors. Snubbing criticism by not addressing it promptly runs counter to our fundamental ethos as scientists, and threatens to erode society's trust in the scientific community.

Sophien Kamoun, Cyril Zipfel
The Sainsbury Laboratory, Norwich, UK.

sophien.kamoun@tsl.ac.uk

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the guidelines at <http://go.nature.com/cmchno>.

Metastability and no criticality

ARISING FROM J. C. Palmer *et al.* *Nature* **510**, 385–388 (2014); <http://dx.doi.org/10.1038/nature13405>

Palmer *et al.*¹ support the idea² of two distinct liquid phases and a low temperature critical point in supercooled water. They do so claiming that molecular simulation of one particular model reveals a stable interface separating two metastable liquids. Here we note that fundamental considerations contradict the idea, and we consider that the data presented do not support the claim. There is a Reply to this Brief Communication Arising by Palmer, J. C. *et al.* *Nature* **531**, <http://dx.doi.org/10.1038/nature16540> (2016).

Binder observes³ that two-liquid criticality defined in terms of a divergent length scale is impossible at supercooled conditions: in the vicinity of a presumed critical point, growing lengths must coincide with growing equilibration times, but the time available to equilibrate can be no longer than the time it takes the metastable liquid to crystallize. Thus, metastability (or instability) implies an upper bound to the size of fluctuations that can relax in the liquid. For water, this size seems to be no larger than 2 or 3 nm, corresponding to volumes containing fewer than 1,000 molecules (see Methods).

Metastable fluctuations on smaller length scales might seem interpretable in terms of a liquid–liquid transition, but the length-scale bound implies it is impossible to know if the interpretation is correct. Further, the interpretation seems unnecessary, because reasonable molecular models known to not exhibit two-liquid behaviour account for equilibrium anomalies of water^{4–6} and non-equilibrium amorphous ices⁷.

Significant fluctuations occur in supercooled water owing to coarsening of ice and competing effects of dynamic heterogeneity. Figure 1 refers to experimental^{8,9} and theoretical^{7,10} information about these behaviours. Fluctuations are largest in the vicinity of the stability temperature, T_s , below which nanometre-scale domains of the liquid are no longer even metastable. Relaxation in that regime is slow because T_s is well below the onset temperature of glass-forming dynamics, T_o . With two (or more) irreversible glass phases of different densities, transient mesoscopic domains will appear as precursors in the reversible melt. These non-equilibrium phenomena can be confused with two-liquid criticality, as illustrated and analysed in ref. 12.

All estimated locations for a critical point in supercooled water are spread over a range of pressures below 1 kbar, and temperatures $T_s < T < T_h$ (ref. 8), arrived at through extrapolations from measurements made well outside that region. Here, T_h is the homogeneous nucleation temperature, below which ice forms rapidly. The one experiment to venture below T_h (ref. 9) finds the liquid persisting for only 10^{-3} s at 227 K, and without a hint of critical fluctuations. Thus, the putative critical temperature would need to be even lower¹¹, were it to exist.

The claim of Palmer *et al.*¹ that numerical simulation demonstrates two-liquid behaviour that is close to criticality and can be scaled to large sizes has, in our view, several technical problems^{12,13}. The behaviour that they report, already reproduced as the result of limiting relaxation of fluctuations¹², is transient and disappears as fluctuations are allowed to relax¹². The issue is not in the reliability of simulation algorithms and codes, but rather in the using of codes in ways consistent with reversibility, which can be challenging owing to slow relaxation. LAMMPS codes used in refs 5 and 12 are standard and documented¹⁴, with scripts freely available upon request, and applications taking care of reversibility^{5,12} establish consistent behaviour among several different models of water.

Technical problems aside, the interpretation by Palmer *et al.*¹ is based upon system sizes too small to demonstrate interfacial scaling,

and their claim of showing a stable interface contradicts some earlier work^{12,15}. Indeed their data are for systems containing only 200 to 600 molecules, and the data can be equally well interpreted in terms of system-size dependence of finite transient domains, not a macroscopic interface separating two phases¹⁶.

Two-liquid coexistence and criticality are in general possible, but not in water where this behaviour would be required to exist at deeply supercooled conditions. Given that fact and that suitable models of liquid water do not exhibit two-liquid coexistence, it seems most fruitful to treat supercooled water on its own terms, as a metastable or unstable non-equilibrium material with the largest fluctuations manifesting

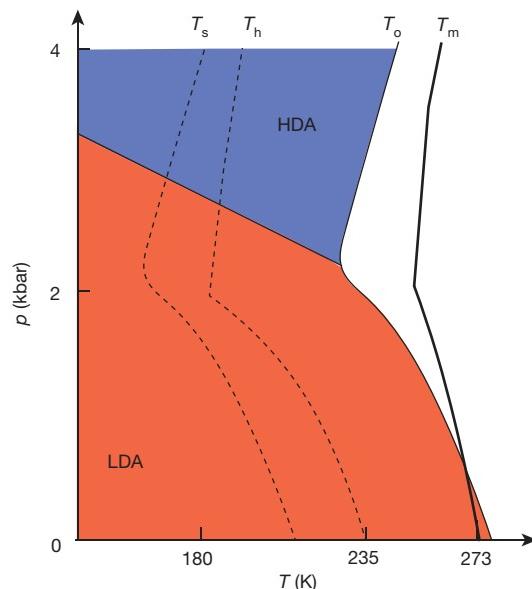


Figure 1 | Phase diagram of supercooled water, with T and p denoting temperature and pressure, respectively. Corresponding states for the simulation model used by Palmer *et al.*¹ are shifted to higher temperatures by 10%–15%. For $T > T_m$, liquid is stable; for $T < T_m$, where ices are stable, irreversible behaviours of the liquid are varied, depending upon experimental protocols. For $T < T_h$, the lifetime of the liquid is of the order of $10^{1\pm 1}$ s or less. This region is sometimes called the ‘no man’s land’ of liquid water because observation of the liquid is difficult for $T < T_h$ (T_h is given in ref. 8). At yet lower temperatures, $T \lesssim T_s$, coarsening rather than simple nucleation becomes rate determining. At that stage, the liquid is dominated by fluctuations, and its lifetime increases with decreasing temperature. Below the onset temperature, T_o , dynamics in the liquid are heterogeneous and intermittent, and far enough below, water can be driven out of equilibrium into high-density and low-density amorphous ices, HDA and LDA. (In this figure, the lines showing T_o and T_s are estimated from theory^{7,10}, and equation (16) of ref. 10 is a formula for the temperature dependence of metastable lifetime. These results have been tested to a limited extent. Further tests await future experiments¹⁶.) The properties and transition temperatures of these glasses depend upon the timescale at which the liquid is driven out of equilibrium¹⁷. The line between HDA and LDA domains marks the p at which the T to reach that timescale is minimum⁷. At very low temperatures, this line relates to a first-order-like non-equilibrium transition between HDA and LDA phases¹⁸. Observations of the transition show a large range of hysteresis with the average of the forward and backward transition pressures being close to that line.

BRIEF COMMUNICATIONS ARISING

ice coarsening. These non-equilibrium phenomena are distinct from equilibrium liquid–liquid criticality.

Methods

At conditions of two-liquid criticality³, the time to equilibrate on length scale ξ is of the order of $\tau_\xi = \tau_R(\xi/a)^3$, where a is a characteristic microscopic length, and τ_R the time to relax the liquid on length scale a . Clearly, $\tau_\xi < \tau_{MS}$, where τ_{MS} is the lifetime of the metastable liquid. Accordingly

$$\xi/a < (\tau_{MS}/\tau_R)^{1/3} \quad (1)$$

wherever criticality might apply. For supercooled water, that regime would be $T \lesssim T_s$, where fluctuations are largest. There, both τ_R and τ_{MS} grow with decreasing temperature T , but estimates of the ratio yield $(\tau_{MS}/\tau_R) \lesssim 10^3$ throughout¹⁰. Thus, because $a \approx 0.2$ or 0.3 nm, $\xi < 2$ or 3 nm. Further details and discussion about uncertainties regarding these estimates are presented elsewhere¹⁶. I note that applying equation (1) at 227 K, where $\tau_{MS} \approx 10^{-3}$ s yields a much larger value for ξ , is not appropriate because experiments at that temperature show water ice nucleation⁹, not large fluctuations or criticality.

D. Chandler¹

¹Department of Chemistry, University of California, Berkeley, California 94720, USA.
email: chandler@berkeley.edu

Received 9 August 2014; accepted 2 December 2015.

- Palmer, J. C. *et al.* Metastable liquid–liquid transition in a molecular model of water. *Nature* **510**, 385–388 (2014).
- Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. Phase behaviour of metastable water. *Nature* **360**, 324–328 (1992).
- Binder, K. Simulations clarify when supercooled water freezes into glassy structures. *Proc. Natl Acad. Sci. USA* **111**, 9374–9375 (2014).

- Holten, V., Limmer, D. T., Molinero, V. & Anisimov, M. A. Nature of the anomalies in the supercooled liquid state of the mW model of water. *J. Chem. Phys.* **138**, 174501 (2013).
- Limmer, D. T. & Chandler, D. The putative liquid–liquid transition is a liquid–solid transition in atomistic models of water. *J. Chem. Phys.* **135**, 134503 (2011).
- Overduin, S. D. & Patey, G. N. An analysis of fluctuations in supercooled TIP4P/2005 water. *J. Chem. Phys.* **138**, 184502 (2013).
- Limmer, D. T. & Chandler, D. Theory of amorphous ices. *Proc. Natl Acad. Sci. USA* **111**, 9413–9417 (2014).
- Holten, V., Bertrand, C. E., Anisimov, M. A. & Sengers, J. V. Thermodynamics of supercooled water. *J. Chem. Phys.* **136**, 094507 (2012).
- Sellberg, J. A. *et al.* Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature. *Nature* **510**, 381–384 (2014).
- Limmer, D. T. & Chandler, D. Corresponding states for mesostructure and dynamics of supercooled water. *Faraday Discuss.* **167**, 485–498 (2013).
- Soper, A. K. Supercooled water: continuous trends. *Nature Mater.* **13**, 671–673 (2014).
- Limmer, D. T. & Chandler, D. The putative liquid–liquid transition is a liquid–solid transition in atomistic models of water. II. *J. Chem. Phys.* **138**, 214504 (2013).
- Chandler, D. Illusions of phase coexistence. Preprint at <http://arXiv.org/abs/1407.6854> (2014).
- Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
- English, N. J., Kusalik, P. G. & Tse, J. S. Density equalisation in supercooled high- and low-density water mixtures. *J. Chem. Phys.* **139**, 084508 (2013).
- Limmer, D. T. & Chandler, D. Time scales of supercooled water and implications for reversible polyamorphism. *Mol. Phys.* **113**, 2799–2804 (2015).
- Limmer, D. T. The length and time scales of water's glass transitions. *J. Chem. Phys.* **140**, 214509 (2014).
- Mishima, O., Calvert, L. D. & Whalley, E. An apparently first-order transition between two amorphous phases of ice induced by pressure. *Nature* **314**, 76–78 (1985).

Competing Financial Interests Declared none.

doi:10.1038/nature16539

Palmer *et al.* reply

REPLYING TO D. Chandler *Nature* **531**, <http://dx.doi.org/10.1038/nature16539> (2016)

We reported¹ for the ST2 model of water advanced free energy calculations using six sampling techniques, all of which show the existence of a low-density liquid (LDL) as well as a high-density liquid (HDL) and a liquid–liquid phase transition (LLPT) between them. In the accompanying Comment², Chandler contends that fundamental arguments³ preclude an LLPT in water and reiterates his claim⁴ that the LDL phase is an artefact associated with poor equilibration.

We point out that although the fundamental argument³ concerns the question of whether critical fluctuations can be detected in metastable systems despite nucleation of the stable phase, it was explicitly stated³ that it has firm implications only for the detection of a critical point, but does not preclude liquid–liquid phase separation. When applying this argument, Chandler² concludes that critical fluctuations larger than 2–3 nm cannot be equilibrated in deeply supercooled water. Following the same analysis but using different values for relevant timescales ($\tau_R \approx 10^{-10}$ s from experimentally derived correlations⁵ and $\tau_{MS} \approx 10^{-3}$ s from experiment⁶, as defined in ref. 2), we estimate that critical fluctuations at 229 K can reach ~ 100 nm—potentially large enough to characterize experimentally.

Regarding putative artefacts arising due to poor equilibration, the LDL persisted in our simulations^{1,7} after relaxing all accessible fluctuations by sampling reversibly between the liquid and crystal regions, and two liquid basins were obtained independently of the sampling method and duration. The LDL basin did not disappear over time

when sampling to and from the crystal, as had been predicted when incorrectly⁴ assuming our calculations to be poorly equilibrated. The free energy exhibits scaling consistent with an LLPT over the range of system sizes that can be explored computationally^{1,7}. Each of these facts is inconsistent with poor equilibration. Moreover, the salient features of our free energy calculations have been reproduced by others^{8,9}, and our code has been publicly available since 2014 (<http://pablonet.princeton.edu/pgd/html/links.html>). The recent demonstration⁹ that adjustment of a single model parameter in ST2 (the hydrogen-bond angular flexibility) makes the LLPT thermodynamically stable with respect to ice Ih/Ic disproves the claim⁴ that crystallization was mistaken for an LLPT.

The main issue is the irreconcilable difference between seemingly identical free energy calculations for the same water model: these identified either two liquids and a crystal as we reported¹, or only one liquid and one crystal⁴. Chandler argues² that LLPT-like artefacts arise from limiting relaxation of fluctuations, but this was only observed when transforming simulation data using a theory whose key assumption is that density fluctuations in HDL decay much faster than bond-orientational fluctuations⁴. In contrast, molecular dynamics simulations show that density is the slowly relaxing variable in the HDL region¹⁰. Chandler's explanation is therefore contradicted by the reversible phase behaviour^{1,8,9} and equilibrium dynamics^{1,10} of ST2. Ultimately, we are confident that continued scrutiny of codes

BRIEF COMMUNICATIONS ARISING

and methods used in the free energy calculations will reveal the cause of the different behaviours predicted for ST2. The question of which one occurs in real water must await an answer by experiment, not by theory or simulation.

Jeremy C. Palmer¹†, Fausto Martelli², Yang Liu¹†, Roberto Car², Athanassios Z. Panagiotopoulos¹ & Pablo G. Debenedetti¹

¹Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA.

email: pdebene@princeton.edu

²Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA.

†Present addresses: Department of Chemical and Biomolecular Engineering, University of Houston, Houston, Texas 77204, USA (J.C.P.); Air Products and Chemicals, Inc., Allentown, Pennsylvania 18195, USA (Y.L.).

1. Palmer, J. C. et al. Metastable liquid–liquid transition in a molecular model of water. *Nature* **510**, 385–388 (2014).
2. Chandler, D. Metastability and no criticality. *Nature* **531**, <http://dx.doi.org/10.1038/nature16539> (2016).

3. Binder, K. Simulations clarify when supercooled water freezes into glassy structures. *Proc. Natl Acad. Sci. USA* **111**, 9374–9375 (2014).
4. Limmer, D. T. & Chandler, D. The putative liquid–liquid transition is a liquid–solid transition in atomistic models of water. II. *J. Chem. Phys.* **138**, 214504 (2013).
5. Torre, R., Bartolini, P. & Righini, R. Structural relaxation in supercooled water by time-resolved spectroscopy. *Nature* **428**, 296–299 (2004).
6. Sellberg, J. A. et al. Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature. *Nature* **510**, 381–384 (2014).
7. Palmer, J. C., Debenedetti, P. G., Car, R. & Panagiotopoulos, A. Z. Response to Comment [arXiv:1407.6854] on Palmer et al., *Nature*, 510, 385, 2014. Preprint at <http://arXiv.org/abs/1407.7884> (2014).
8. Poole, P. H., Bowles, R. K., Saika-Voivod, I. & Sciortino, F. Free energy surface of ST2 water near the liquid–liquid phase transition. *J. Chem. Phys.* **138**, 034505 (2013).
9. Smallenburg, F. & Sciortino, F. Tuning the liquid–liquid transition by modulating the hydrogen-bond angular flexibility in a model for water. *Phys. Rev. Lett.* **115**, 015701 (2015).
10. Palmer, J. C., Martelli, F. & Debenedetti, P. G. Density and bond-orientational relaxations in supercooled tetrahedral liquids. Preprint at <http://arXiv.org/abs/1511.01959> (2015).

doi:10.1038/nature16540

MOLECULAR EVOLUTION

Sex accelerates adaptation

An analysis confirms the long-standing theory that sex increases the rate of adaptive evolution by accelerating the speed at which beneficial mutations sweep through sexual, as opposed to asexual, populations. SEE LETTER P.233

MATTHEW R. GODDARD

When compared with the asexual alternative of simple cloning, sex seems like a complicated way of reproducing. The need for fast and efficient reproduction lies at the heart of Darwinian natural selection, so why sex exists is a conundrum that has fascinated biologists for more than 100 years¹. In this issue, McDonald *et al.*² (page 233) directly confirm the long-held theory that the advantage of sex lies in its ability to expose individual mutations to the actions of natural selection.

Sex involves the shuffling (recombination) of chromosomes from parents, followed by the separation of these newly assorted chromosomes into reproductive cells called gametes, which then fuse through mating. As well as being more complicated than asexual reproduction, this mechanism risks breaking apart collections of genes that have proved to be useful. In animals, only females can give birth, and mate finding and courtship impose further challenges. Given these disadvantages, it is not immediately clear why sexual reproduction has persisted.

Some of the mutations that accrue in genomes over time affect an organism's ability to reproduce and compete for resources (fitness). The net fitness of an individual is the sum of these accrued mutations. Conventional theories^{3,4} suggest that selection in asexually reproducing populations is affected only by this net genomic fitness.

In this scenario, when a positive mutation arises in a genome that already harbours negative mutations, the negative mutations might overwhelm the positive one, leading to the removal of the whole genome from the population by natural selection and the loss of the positive mutation. However, if a positive mutation confers a strong-enough fitness benefit to outweigh the combined value of the negative mutations, then the genome is likely to become more common over generations — possibly becoming a permanent part of (fixed in) the population owing to positive selection. Negative mutations become common by hitch-hiking with positive ones, and thus restrict population fitness. In summary, individual mutations in asexual populations may be masked from the actions of selection, because they are entangled in genomes.

Sexual populations theoretically do not have this problem^{3,4}. Recombination and the random partitioning of chromosomes allow positive mutations to become dissociated from negative ones. By analogy, sex allows selection to pluck rubies from rubbish⁵. Furthermore, it enables positive mutations that arise in different genomes to be recombined into the same genome, rather than competing with one another as they would in an asexual population⁴. In sexual populations, many positive mutations, mostly free from hitch-hiking mutational rubbish, can become common simultaneously. This is predicted to increase the rate and extent of adaptive evolution¹.

A series of experimental-evolution studies

supports the idea that sex speeds up adaptive evolution^{6–8}. However, much less work has focused on the molecular mechanisms that underpin this advantage. One study⁹ inferred that sex accelerates adaptation by separating positive mutations from negative ones, but did not directly identify the mutations that arose.

McDonald *et al.*, however, have done just that. First, they caused sexual and asexual populations of yeast (*Saccharomyces cerevisiae*) to evolve for approximately 1,000 generations in a simple laboratory environment, to which the sexual populations adapted more rapidly. Then, building on previous studies, the authors used DNA-sequencing approaches to dissect and track the various single DNA-base mutations that arose, evaluating populations at regular time points during evolution.

A similar range of mutations initially arose in all populations, some of which affected protein function, with others having no effect. The authors reasonably assumed that only those that altered protein function would affect fitness. In asexual populations, the different types of mutation all had roughly the same chance of eventually becoming fixed, indicating that selection could not discriminate between individual mutations. Fewer mutations became fixed in sexual populations. Those that did tended to alter protein function, and thus also, presumably, fitness. This observation suggests that sex improved the efficiency with which selection acted on individual mutations.

To directly test the effects of specific mutations on fitness, McDonald *et al.*

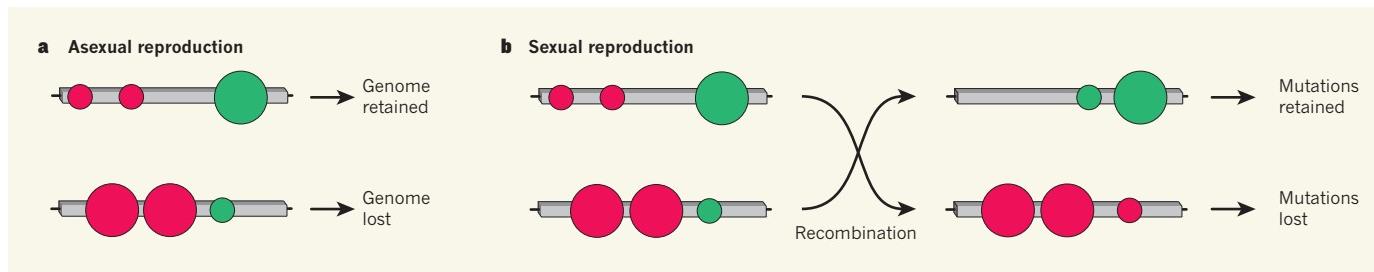


Figure 1 | Picking rubies from the rubbish. Over time, genomes accrue mutations that have either a positive (green) or a negative (red) effect on an individual's fitness (in this simple schematic, the relative benefit or cost of each mutation is indicated by size). McDonald *et al.*² compared how selection acts on mutations in asexual and sexual populations. **a**, During asexual reproduction, selection occurs on the basis of overall genome

fitness. Positive mutations may be removed from the population, and negative mutations can hitch-hike along with a positive one of greater value. **b**, During sexual reproduction, chromosomes are shuffled by recombination, changing the mutations that are grouped together in offspring. This process enables individual mutations to be independently retained or removed by selection.

conducted mini-evolution experiments and tracked the change in frequency of individual mutations in the population. This key step revealed that groups of positive and negative mutations remained together in asexual populations. These groups competed with one another — some became common over generations, meaning that negative mutations persisted by hitch-hiking. By contrast, recombination meant that no groups of mutations persisted in sexual populations, and negative mutations did not become common.

These comprehensive experiments provide the long-awaited confirmation that sex accelerates adaptation by sorting the beneficial from the deleterious. Sex shuffles mutations between genomes, enabling natural selection to act on individual mutations more efficiently (Fig. 1). Selection is comparatively blind in asexual populations, because the effects of individual mutations are consistently hidden in genomes.

But McDonald and colleagues' study leaves several aspects of sexual reproduction still to be clarified. First, the authors primarily examined changes of single DNA bases. However, mutations that duplicate, remove or rearrange whole segments of DNA are also important for adaptation. As the authors acknowledge, the effect of sex on these mutations remains to be evaluated.

Second, the study used yeast that has one copy of each chromosome, whereas most sexual species have two copies, and natural selection works slightly differently when there are two chromosomes. Third, most species inhabit complex environments, which have a variety of selection pressures whose strength varies over space and time. Although the current study elegantly shows how sex provides advantages during adaptation to simple environments, it is not clear how this translates to more-complex ones. Some work suggests that sex can also accelerate adaptation to complex environments¹⁰; however, the underlying molecular mechanisms are not known.

Finally, we do not yet know why sex arose in the first place. One theory suggests that parasitic genetic elements, which persist in genomes despite conferring no fitness benefit, might promote cell fusion and recombination¹¹. Few experiments have tested this theory, however¹². It might well be that the evolution of sex was driven by completely different forces from those — neatly defined by McDonald *et al.* — that we now know maintain it. ■

Matthew R. Goddard is at the School of Life Sciences, University of Lincoln, Lincoln LN6 7TS, UK, and at the School of Biological Sciences, University of Auckland, New Zealand.
e-mail: mgoddard@lincoln.ac.uk

- Burt, A. *Evolution* **54**, 337–351 (2000).
- McDonald, M. J., Rice, D. P. & Desai, M. M. *Nature* **531**, 233–236 (2016).

- Fisher, R. A. *The Genetical Theory of Natural Selection* (HardPress, 2013).
- Muller, H. J. *Am. Nat.* **66**, 118–138 (1932).
- Peck, J. R. *Genetics* **137**, 597–606 (1994).
- Colegrave, N. *Nature* **420**, 664–666 (2002).
- Goddard, M. R., Godfray, H. C. J. & Burt, A. *Nature* **434**, 636–640 (2005).
- Moran, L. T., Schmidt, O. G., Gelarden, I. A., Parrish, R. C. 2nd & Lively, C. M. *Science* **333**, 216–218 (2011).
- Gray, J. C. & Goddard, M. R. *BMC Evol. Biol.* **12**, 43 (2012).
- Gray, J. C. & Goddard, M. R. *Ecol. Lett.* **15**, 955–962 (2012).
- Hickey, D. A. & Rose, M. R. in *The Evolution of Sex: An Examination of Current Ideas* (eds Michod, R. E. & Levin, B. R.) 161–193 (Sinauer, 1988).
- Giraldo-Perez, P. & Goddard, M. R. *A Proc. R. Soc. B* **280**, 20131875 (2013).

This article was published online on 24 February 2016.

SUPERCONDUCTORS

Exponential boost for quantum information

Quantum computers will one day wildly outperform conventional machines. An experimental feat reveals a fundamental property of exotic superconductors that brings this quantum technology a step closer. SEE LETTER P.206

JASON ALICEA

Quantum computers promise a technological revolution that will easily surmount otherwise impenetrable problems in cryptography, quantum simulation, drug design and more. Building the hardware has been challenging, however, because unavoidable random noise from the environment readily corrupts quantum bits (qubits). On page 206 of this issue, Albrecht *et al.*¹ pursue an elegant strategy for sidestepping this obstruction. They discover a key property of ‘Majorana modes’ in superconducting wires that can be used to engineer qubits that are immune to noise by default.

A simple analogy conveys the basic notion of Majorana modes. Imagine a line of schoolchildren, each holding hands with their neighbours, leaving an uncoupled free hand at either end of the chain (Fig. 1a). The electrons in certain exotic superconducting wires, which physicists are becoming highly adept at building, entangle in an analogous pattern²: half of each electron couples with its rightward neighbour and the other half couples with its leftward neighbour. Majorana modes are the leftover ‘free hands’, or unpaired electron halves, at the superconductor’s ends. Roughly speaking, an electron has been cut in two, and the fragments are separated across the wire.

Together, the two Majorana modes form a single quantum level that can be empty or filled by an electron (Fig. 1b). Theory predicts² that the energy required to populate that state decreases exponentially as the distance between Majorana modes increases. At the extreme limit, at which the energy is exactly zero, it becomes impossible to detect this level’s occupation by performing a local measurement at the wire’s ends — or elsewhere, for that matter. The individual Majorana modes carry neither energy nor any other locally detectable

property that could unveil the precise quantum state formed with its distant partner. Instead, that information spreads out globally across the system, securely hidden from ordinarily problematic noise sources.

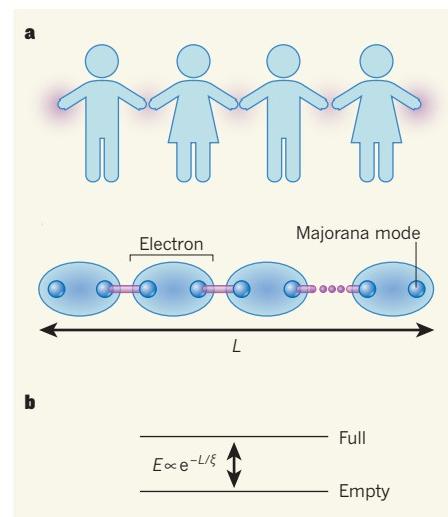


Figure 1 | Majorana modes in superconducting wires. **a**, The linked hands in a chain of children mimic the entanglement of electrons in specially constructed superconducting wires²: half of each electron couples to its rightward neighbour, with the other half coupling to its leftward neighbour (purple dots indicate additional electrons that are similarly entangled). The free hands at the ends are analogous to Majorana modes (unpaired electron halves, separated by distance L) in the superconductor. **b**, These two Majorana modes form a single quantum level that can be either empty or filled by an electron. The equation represents the predicted relationship between the energy, E , needed to fill that level and L ; ξ is an exponential decay constant. Albrecht *et al.*¹ have confirmed the exponential suppression of E as the wire length increases.

Even more interesting multi-wire circuits enable the user to process quantum information in exquisitely precise ways, also largely immune to noise, by manoeuvring Majorana modes around one another — akin to braiding strands of hair (see ref. 3 for one promising hardware blueprint). Majorana modes are therefore widely coveted vaults for quantum information. So where do we stand regarding the implementation of noise-resistant, Majorana-based quantum computation?

Pioneering experiments^{4,5} have made headlines for detecting credible Majorana-mode signatures using measurements of electrical conductance in superconducting devices. (The existence of Majorana modes can be inferred from local probes, but the quantum information that they encode cannot.) Albrecht and colleagues break experimental ground by quantifying how Majorana modes evolve as they are pulled apart. The authors studied ultra-high-quality superconducting wires with lengths ranging from 330 to 1,500 nanometres, using a clever measurement scheme to determine how much energy it costs to add just one electron to the superconductor.

If Majorana modes indeed form in the authors' devices, the energy cost should decrease exponentially on moving to progressively longer wires. This is precisely what the researchers detect — uncovering a fundamental Majorana feature that is intimately related to the devices' applicability to quantum computing. The measured exponential decay constant turns out to be surprisingly short (about 250 nm), indicating that even modestly sized systems might harbour nearly ideal Majorana modes, and correspondingly ideal qubits.

Various other nuances of the length dependence for the measured energies also agree well with theoretical expectations⁶. Collectively, the data accumulated in this latest experiment seem exceedingly hard to explain using conventional physics. The results, however, offer more than just additional evidence for the existence of Majorana modes in superconducting wires: the unprecedented characterization primes the field for a fascinating new era of Majorana control.

In particular, the stage now seems set for quantitative tests of the basic principles that underlie intrinsically fault-tolerant quantum information processing. Two crucial experimental challenges for this endeavour are to develop techniques for dynamically manipulating Majorana modes — that is, to create, transport and fuse them within a single device — and to demonstrate successful readout of the hidden information encoded through their quantum states. These capabilities will, in turn, enable a wide range of experiments, even in surprisingly simple devices, that inch towards applications.

Future research should aim to quantify the protection of quantum information stored in a prototype Majorana qubit, and to meaningfully

contrast its behaviour with that of conventional qubits. Braiding Majorana modes to implement fault-tolerant information processing poses another seminal challenge for the field. Proof-of-principle demonstrations of these concepts could pave the way for a new generation of robust, scalable quantum-computing hardware, while offering fascinating glimpses into previously unobserved facets of quantum mechanics and a host of surprises along the way. ■

Jason Alicea is in the Institute for Quantum Information and Matter, and in the Division of

Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA.
e-mail: aliceaj@caltech.edu

- Albrecht, S. M. et al. *Nature* **531**, 206–209 (2016).
- Kitaev, A. Yu. *Sov. Phys. Usp.* **44** (suppl.), 131–136 (2001).
- Hyart, T. et al. *Phys. Rev. B* **88**, 035121 (2013).
- Mourik, V. et al. *Science* **336**, 1003–1007 (2012).
- Nadj-Perge, S. et al. *Science* **346**, 602–607 (2014).
- Stanescu, T. D., Lutchyn, R. M. & Das Sarma, S. *Phys. Rev. B* **87**, 094518 (2013).

PLANT BIOLOGY

LURE is bait for multiple receptors

In flowering plants, sperm-containing pollen tubes are guided towards ovules by attractants from the female reproductive organ. Receptors for the attractant molecule AtLURE1 have now been found. SEE LETTERS P.241 & P.245

ALICE Y. CHEUNG & HEN-MING WU

For flowering plants to achieve fertilization, pollen must transport sperm across long distances. Sperm-containing pollen grains land on the stigma of the female reproductive organ (the pistil), but the female gametophyte structures that bear eggs are located in distant ovules, so each grain produces a pollen tube that grows towards them¹ (Fig. 1a). How pollen tubes find their target has long puzzled biologists. The female gametophyte is known to produce chemoattractant molecules, such as cysteine-rich peptides called LUREs^{1,2}, but the identity of their receptors on pollen tubes has been unclear. Two papers in this issue^{3,4} identify several molecules on the cell membrane that are involved in sensing one such attractant — AtLURE1 — in the model plant *Arabidopsis thaliana*⁵. These discoveries underscore the molecular complexity of this male–female communication process, and provide a foundation for understanding the mechanism by which pollen tubes sense attractants.

It is well established that pollen-specific receptor-like kinase (RLK) proteins can regulate the growth of pollen tubes⁶. These proteins typically have three domains: an ectodomain that interacts with extracellular signal molecules; a membrane-spanning domain; and a cytoplasmic domain that attaches phosphate groups to target molecules, inducing cellular responses to incoming signals (Fig. 1b). Using different genetic strategies and starting from an overlapping list of almost 30 pollen-expressed RLKs, the two groups searched for

proteins that support ovule targeting by pollen tubes.

On page 241, Wang and colleagues³ report two pairs of closely related RLKs. The authors named the first pair male discoverer 1 (MDIS1) and MDIS2, and the second pair MDIS1-interacting RLK 1 (MIK1) and MIK2. Mutation in the genes that encode each of these four RLKs compromised ovule targeting, and further genetic analysis suggested that MDIS1 and MIK1 act in the same pathway. Next, the authors performed attractant assays in a ‘semi *in vivo*’ system, in which pollen tubes are first allowed to grow through the pistil, which primes them to respond to attractants^{2,5} when subsequently placed under *in vitro* growth conditions. The assay confirmed that mutations in the MDIS1, MIK1 and MIK2 genes impair the ability of pollen tubes to target AtLURE1, although each mutation suppressed targeting only moderately.

Using similar assays, Takeuchi and Higashiyama⁴ (page 245) identified another set of RLK receptors for AtLURE1. One, named pollen-specific receptor kinase 6 (PRK6), was essential for pollen tubes to target AtLURE1 in the semi *in vivo* assay. However, in the pistil, PRK6 mutant pollen tubes displayed only moderate defects in growth and ovule targeting. When the authors combined PRK6 mutations with mutations in the related genes PRK1, PRK3 and PRK8, pollen tubes displayed more-severe guidance defects, including failure to enter ovules.

The attractants identified so far show species specificity^{1,2}. Both Wang *et al.* and Takeuchi and Higashiyama showed that they could

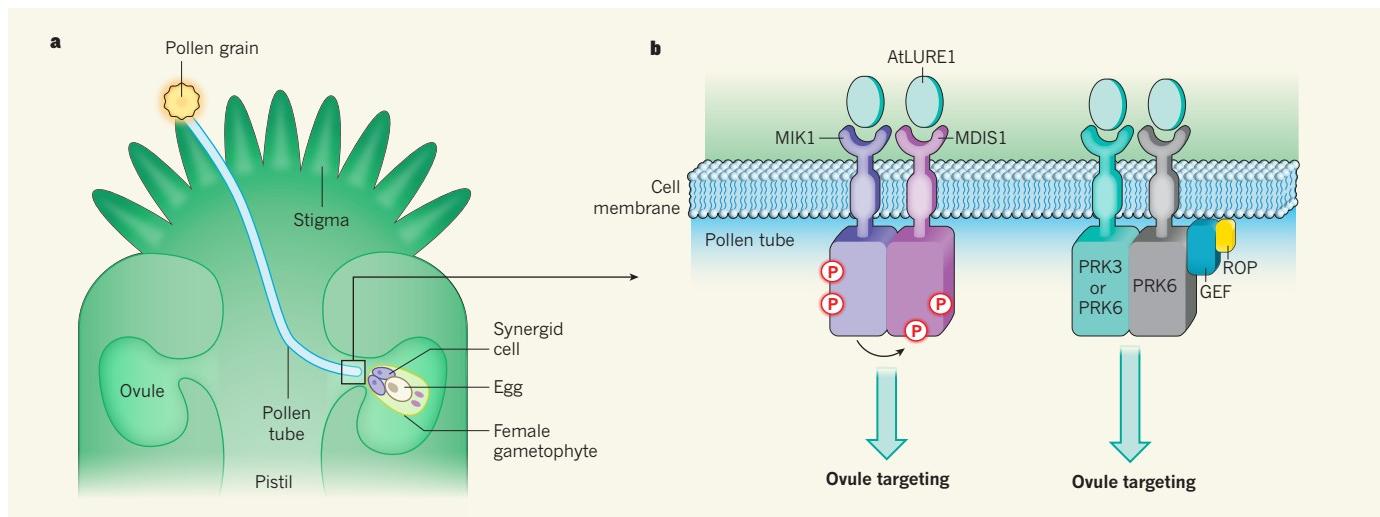


Figure 1 | Long journey for pollen grains. **a**, In flowering plants, sperm-containing pollen grains land on the receptive stigma of the female reproductive organ (the pistil), and form pollen tubes that grow towards egg-bearing structures called female gametophytes, which are located in ovules. Paired synergid cells in the female gametophyte release molecules such as AtLURE1 that attract the tubes. When the pollen tube penetrates a synergid, it ruptures, releasing two sperm for fertilization. **b**, Wang *et al.*³ and Takeuchi and Higashiyama⁴

Higashiyama⁴ have identified receptor-like kinase (RLK) proteins on pollen tubes that are involved in ovule targeting. Wang *et al.* showed that AtLURE1 binds the RLKs MIK1 and MDIS1, promoting their dimerization and inducing MIK1 to add phosphate groups (P) to itself and to MDIS1. Takeuchi and Higashiyama showed that the RLK PRK6 interacts with itself and PRK3, and with guanine-exchange factors (GEFs) that activate Rho GTPase proteins from plants (ROPs), leading to ovule targeting.

significantly enhance the ability of pollen tubes from a relative of *Arabidopsis*, *Capsella rubella*, to target *A. thaliana* AtLURE1, by engineering them to express MDIS1 or PRK6, respectively — experiments that further support the role of these RLKs in attractant sensing. Taken together, the groups' results indicate that the perception system for AtLURE1 involves multiple RLKs that are functionally redundant, acting together to support ovule targeting by pollen tubes and ensure reproductive success.

Wang *et al.* provided biochemical and biophysical data to demonstrate a physical and functional interaction between their two pairs of RLKs, and to show that AtLURE1 affects the RLKs' interaction and binds directly to MDIS1, MIK1 and MIK2 with different affinities. Technical difficulties that arose owing to a lack of binding specificity prevented Takeuchi and Higashiyama from reporting similar AtLURE1–PRK interaction experiments, although Wang *et al.* demonstrated that AtLURE1 did not bind appreciably to PRK3 in a test that they did to demonstrate the specificity of AtLURE1 for their RLKs. These differences might be due to variations in protein preparation and quality, or assay conditions, between the two groups; they will need to be resolved.

Using leaf-cell-based assays, both studies next investigated the mechanisms that mediate AtLURE1 signalling (Fig. 1b). Takeuchi and Higashiyama showed that PRK3 and PRK6 interact with guanine-exchange factors that activate Rho GTPase proteins, affirming a known link between PRK proteins and these signal mediators⁶. How AtLURE1 affects these interactions remains to be shown. Wang *et al.* found that AtLURE1 induces MDIS1–MIK1

binding and promotes phosphorylation of the two RLKs by MIK1, implying that changes in the phosphorylation states of these kinases underlie their ability to transform the attractant signal into a guidance response. Future experiments should confirm these interactions in pollen tubes, and test whether these pathways intersect as segments of the same AtLURE1-triggered cascade.

Finally, both groups showed that the location of their RLKs was altered by AtLURE1, bolstering the assertion that these are bona fide AtLURE1 receptors. Wang *et al.* reported that AtLURE1 induced the removal of MDIS1 from the cell membrane — a change that implies a receptor response to binding. Takeuchi and Higashiyama demonstrated that AtLURE1 altered the distribution of PRK6 around the apex of the pollen tube such that it concentrated on tube surfaces closer to the attractant, correlating receptor localization with a change in growth direction.

RLKs have crucial roles in plant development, reproduction and responses to environmental challenges. These studies now persuasively establish that RLKs are involved in attractant-sensing by pollen tubes. Moreover, they support the idea that functional redundancy between receptors — and between attractants, as previously suggested^{2,5} — is perhaps genetically hardwired to ensure reproductive success.

However, this redundancy raises a perplexing question about how AtLURE1 differentiates between potential targets. To capitalize on redundant receptors, AtLURE1 has apparently evolved to interact with a range of RLKs, even those with other specialized functions. For instance, Wang *et al.* found that AtLURE1

binds PXY, a close relative of MIK1 that controls vascular differentiation⁷, with an affinity comparable to that for MIK1. However, an attractant closely related to AtLURE1 does not seem^{3,8} to interact with an RLK called ERECTA that controls plant architecture and cell shape at the leaf surface. Clearly, there is a need to determine how cysteine-rich peptide attractants such as LUREs identify the receptors capable of mediating ovule targeting. It will also be interesting to investigate the possibility of functional crossover by other pairs of RLK and growth regulators, including PXY and ERECTA and their interaction partners, if they are expressed in regions close to where male–female communication occurs.

The arsenal of signalling molecules in plants — in particular peptide signal molecules⁹ and RLKs^{4,5} — is immense. It will not be surprising if more attractant–receptor pairs are discovered. The current studies, together with our knowledge of other growth regulatory molecules that interact with pollen tubes before they encounter ovule attractants¹⁰, bring us closer to fully understanding a process that is vital for plant reproduction. ■

Alice Y. Cheung and Hen-Ming Wu
are in the Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, Massachusetts 01003, USA.
e-mails: acheung@biochem.umass.edu; hmwu@biochem.umass.edu

- Dresselhaus, T. & Franklin-Tong, N. *Mol. Plant* **6**, 1018–1036 (2013).
- Higashiyama, T. & Takeuchi, H. *Annu. Rev. Plant Biol.* **66**, 393–413 (2015).
- Wang, T. *et al.* *Nature* **531**, 241–244 (2016).

4. Takeuchi, H. & Higashiyama, T. *Nature* **531**, 245–248 (2016).
5. Takeuchi, H. & Higashiyama, T. *PLoS Biol.* **10**, e1001449 (2012).
6. Cheung, A. Y. & Wu, H.-M. *Annu. Rev. Plant Biol.* **59**, 547–572 (2008).
7. Fisher, K. & Turner, S. *Curr. Biol.* **17**, 1061–1066 (2007).
8. Lee, J. S. *et al.* *Nature* **522**, 439–443 (2015).
9. Qu, L.-J., Li, L., Lan, Z. & Dresselhaus, T. *J. Exp. Bot.* **66**, 5139–5150 (2015).
10. Wu, H.-M., Wang, H. & Cheung, A. Y. *Cell* **82**, 395–403 (1995).

SUSTAINABLE CHEMISTRY

Putting carbon dioxide to work

Carbon dioxide is an abundant resource, but difficult for industry to use effectively. A simple reaction might allow it to be used to make commercial products more sustainably than with current processes. SEE LETTER P.215

ERIC J. BECKMAN

As raw materials go, there is a lot to like about carbon dioxide — it is available everywhere, inexpensive, non-flammable and less toxic than most of the chemicals widely used in industrial processes. But it is relatively unreactive, making it difficult to activate so that it can be transformed into desirable compounds. Nevertheless, in nature, many plants have evolved molecular machinery that overcomes the inherent stability of CO₂ to use it to make biological building blocks (sugars) and materials (polysaccharides). Inspired by the carbon–carbon bond-formation processes used by plants, Banerjee and colleagues¹ (page 215) have identified a synthetic route that not only uses CO₂ to make useful compounds, but also involves tractable processing conditions. Their route is simple, is

potentially more sustainable and economical than the one it is designed to replace, and could be applicable to a variety of product types.

CO₂ has been used as a raw material by the chemical industry in the past^{2,3}, albeit rather sparingly, to make urea (a fertilizer and building block for the chemical industry) and cyclic carbonate (a solvent). The processes were commercialized not because they were more sustainable than other routes, but because the chemistry was available to make these valuable products economically. Scientists have been interested in expanding the role of CO₂ as a raw material for many years, but, for the most part, either the compounds generated from it were not sufficiently useful to merit industrial production, or the processes involved were too energy-intensive or inefficient to warrant further development.

The processes that have been successfully

adopted to make commercial products from CO₂ were typically preceded by breakthroughs in chemistry and/or catalysis. For example, groundbreaking work on catalyst design⁴ allowed CO₂ to be polymerized with another compound, propylene oxide, to create polycarbonate polyols — important building blocks for polyurethanes, and saleable products in their own right. This work was scaled up and commercialized by Novomer, a chemistry-technology company in Ithaca, New York; the international chemical company Bayer has also pursued this role of CO₂ using their own catalysts⁵. Banerjee and colleagues now report new chemistry to make another valuable molecular building block from CO₂.

The authors used caesium carbonate, a simple salt, to activate organic substrates that could then be reacted with CO₂. Their key finding is that CO₂ can be reacted with 2-furan carboxylate (FC; Fig. 1) to form furan-2,5-dicarboxylic acid (FDCA). This is notable because FC is readily derived from biomass waste material, such as maize (corn) stover and sawdust. Furthermore, FDCA is one of the monomers used to generate polyethylene furandicarboxylate (PEF) — a plant-based polyester that is being commercialized⁶ to compete with the widely used plastic polyethylene terephthalate (PET), which is derived from petrochemicals.

Banerjee *et al.* show that the caesium carbonate can be recycled, and that the product can be separated easily from the reaction mixture. Both of these features will aid in scaling up the reaction. Production of PEF results in fewer carbon emissions than production of PET (ref. 6), but the authors' route to FDCA should reduce the overall carbon footprint still further. Once scaled up, the new route might be less wasteful — needing fewer raw materials and less energy — than the conventional industrial synthesis of FDCA, which uses fructose as a starting material.

Synthetic processes involving CO₂ as a raw material can be considered more sustainable than existing processes only if the chemistry involved reduces environmental impacts over the entire life cycle of the process. Carbon footprint is only one of several metrics⁷ used to gauge the environmental impact of a product; other considerations include the potential to increase acidification (acid rain) or to trigger photochemical oxidation (smog). Even though Banerjee and co-workers' process seems to be much less wasteful than the fructose route to FDCA, a comparison of the life-cycle impacts of the two routes will need to be performed to ensure that it is truly more sustainable.

The authors also show that benzene can be reacted with CO₂ and caesium carbonate to form benzoic acid in a single step (see Fig. 3c of the paper¹). This is intriguing because it has been known⁸ since the 1950s that benzoic acid can be transformed into terephthalic acid, one

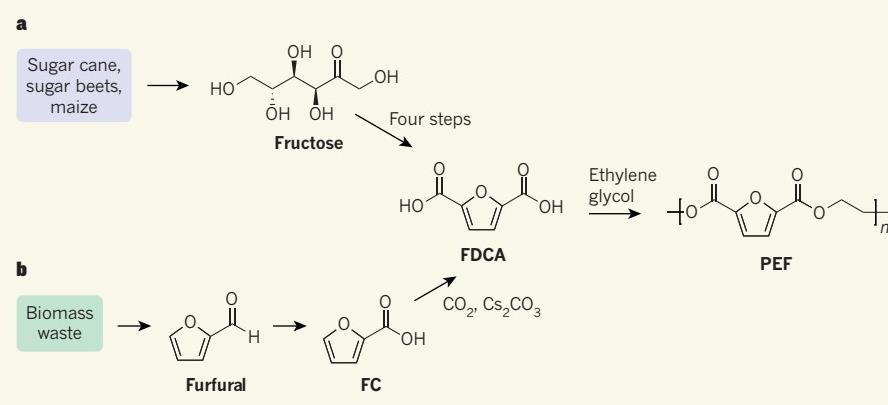


Figure 1 | Synthetic routes to polyethylene furandicarboxylate (PEF). **a**, The polymer PEF is being commercialized as a sustainable alternative to polyethylene terephthalate, a widely used plastic. In the conventional route to PEF, fructose derived from plants is converted by way of a four-step process⁸ to furan-2,5-dicarboxylic acid (FDCA), which can be reacted with ethylene glycol to make PEF. **b**, Banerjee *et al.*¹ report that FDCA can also be made by reacting 2-furan carboxylate (FC) with carbon dioxide in the presence of caesium carbonate (Cs₂CO₃). The reaction could form part of a synthetic route to PEF that is more sustainable than that detailed in **a**. In the new route, biomass waste is first converted to furfural, which is oxidized to make FC.

of the monomers used to make PET. Although the initial reaction yields reported by Banerjee and colleagues are low, the finding raises the possibility that PET, like PEF, could be made using CO₂. If the yields can be improved, then this chemistry would be a marked improvement on the current multistep route used by industry to make terephthalic acid.

More than 45 million tonnes of PET are produced annually⁹, making it one of the largest potential synthetic 'sinks' for CO₂. That said, no synthesis that uses CO₂ will lead to sizeable reductions in atmospheric concentrations of the gas. Nevertheless, finding sustainable

uses for abundant resources such as CO₂ as alternatives to non-renewable resources remains a worthy goal. More broadly, Banerjee and co-workers' results suggest that the molecular machinery devised by chemists will follow the example of plants, by evolving to use CO₂ efficiently to create the feedstocks and materials that we need. ■

Eric J. Beckman is in the Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA.
e-mail: beckman@pitt.edu

- Banerjee, A., Dick, G. R., Yoshino, T. & Kanan, M. W. *Nature* **531**, 215–219 (2016).
- Mavrovic, I., Shirley, A. R. & Coleman, G. R. in *Kirk-Othmer Encyclopedia of Chemical Technology* <http://doi.org/bc6vx9> (Wiley, 2010).
- Busch, H.-J. in *Ullmann's Encyclopedia of Industrial Chemistry* 6th edn, 427–455 (Wiley, 2003).
- Coates, G. W. & Moore, D. R. *Angew. Chem. Int. Edn* **43**, 6618–6639 (2004).
- Haider, K. W., McDaniel, K. G., Hayes, J. E. & Shen, J. US Patent 7,977,501 (2011).
- Eerhart, A. J. J. E., Faaij, A. P. C. & Patel, M. K. *Energy Environ. Sci.* **5**, 6407–6422 (2012).
- Bare, J. C. *J. Ind. Ecol.* **6**, 49–78 (2002).
- Raecke, B., Stein, W. & Schirp, H. US Patent 2,794,830 (1957).
- Sheldon, R. A. *Green Chem.* **16**, 950–963 (2014).

ECOLOGY

Vegetation's responses to climate variability

Satellite data have allowed scientists to generate a quantitative model to assess the response rates of different ecosystems to climate variability. The index provides a tool for comparing regional sensitivity and resilience. SEE LETTER P.229

ALFREDO HUETE

Space offers a unique vantage point from which to investigate the sensitivity of Earth's ecosystems to climate variability. In this issue, Seddon *et al.*¹ (page 229) use 14 years of satellite observations from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) to obtain monthly data on the responses of vegetation to variability in water availability, cloudiness and air temperature. The researchers then applied a new empirical tool — the vegetation sensitivity index — to identify ecologically sensitive areas that exhibit either amplified or slowed responses to climate variability in comparison to other regions. They use their findings to begin to explore why some regions seem to be more vulnerable than others.

Current ecological theory states that, as ecosystems approach critical thresholds (also referred to as tipping points), they become unstable and respond more acutely to external perturbations². Knowledge of these thresholds is key to the sustainable management of ecosystems and to anticipating irreversible changes and/or ecological collapse. But predicting where and when such transitions will occur remains a challenge.

Studies of ecosystem resilience generally monitor trends in productivity or biodiversity in relation to changes in mean climate states, rather than in response to climate variation. A widely used metric of ecological resilience relates changes in the productivity of vegetation to variations in annual rainfall: this is

known as rainfall-use efficiency (RUE), and is often applied to dryland areas^{3,4}. RUE facilitates cross-biome comparisons and establishes a hypothetical threshold beyond which ecological transitions or ecosystem collapse may occur. Seddon *et al.* have greatly extended such sensitivity assessments by including variability in several climate drivers — air temperature, water availability and cloud level — and relating these to vegetation productivity on monthly timescales.

The authors' vegetation sensitivity index (VSI) allows them to assess vegetation-productivity responses to climate variability across all terrestrial ecosystems, from tropical forests and temperate grasslands to the Arctic tundra and alpine areas. The VSI quantifies patterns and drivers of ecological sensitivity by identifying ecologically vulnerable areas, and includes a 'weighting' for the various climate factors that contribute to ecological change.

Like other global assessments of vegetation growth in response to climatic factors⁵, the researchers' index identified amplified vegetation responses to climate variability in areas that are experiencing rapid warming, such as the Arctic tundra and alpine regions (Fig. 1). Tropical forests around the globe also showed amplified sensitivity, which was associated with cloud and light variations. However, in contrast to earlier work⁵, the authors found water availability to be an important driver of sensitivity in tropical forests in central Africa. This is in agreement with the results of a recent study⁶ that also reported water to be a major driver of vegetation productivity in



Figure 1 | Vulnerable to variability. Seddon *et al.*¹ identify amplified vegetation responses to climate variability in areas that are experiencing rapid warming, such as the Tibetan plateau.

African tropical forests. In addition, Seddon *et al.* found grasslands to be the ecosystem that is most sensitive to water availability, in accordance with other work⁷.

It is worth noting that Seddon and colleagues' work derived the relative influence of different climate variables on productivity solely on the basis of satellite observations, and without making assumptions based on hypothesized ecological tolerance limits, as in previous work⁵.

Identifying ecologically sensitive areas with amplified or slower climate response rates is a valuable first step for identifying regions of pending ecological change and for developing sustainable management practices. Satellite data can reveal useful information on vegetation dynamics and provide promising opportunities to measure ecosystem changes and responses to climate variability, as Seddon *et al.* demonstrate. However, satellite information is only one piece of the larger picture, and needs to be integrated with information collected on the ground if we are to fully understand the functional properties of vegetation communities and the plant physiological mechanisms that contribute to ecosystem resilience in the face of climate change⁸.

For example, the functional and structural properties of a young forest may differ from those of a mature forest, and the two systems may therefore respond differently to the same climate variability. Furthermore, other ecosystem parameters may be better suited to assessing responses to climate variability. Intrinsic functional parameters such as efficiency of water and light use are also worthy of further attention⁸, because these may better represent an ecosystem's photosynthetic capacity.

The role of biodiversity in driving differences in ecological sensitivity also needs further exploration. Although biodiversity cannot be directly assessed with the coarse resolution of MODIS satellite data, broad-scale losses in species diversity could easily amplify ecosystem responses to extreme climate events and ecosystem disturbance, and result in lower ecological resilience⁹. These amplified responses may be measurable with satellite indices such as the VSI, although the impact of changes in biodiversity may not be known without *in situ* data. Low species diversity may also prolong ecosystem recovery after major disturbances.

Although Seddon *et al.* addressed lag and memory effects, in which an ecosystem's response depends on both current and past climate conditions¹⁰, a greater insight into the relationship between such effects and measurements of sensitivity is needed. And there remains an overall lack of understanding of the complex interactions between climate events and ecosystem responses across various temporal scales¹¹. However, the authors' findings highlight the necessity of understanding basic ecological sensitivity and recognizing areas that

are vulnerable to climate variability, especially in a warming climate. Only through an understanding of vegetation's responses to current climate variability can we improve predictions of the future consequences of such variability on our planet's ecosystems and biodiversity, as well as on our own food security and welfare. ■

Alfredo Huete is in the Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, Ultimo, New South Wales 2007, Australia.
e-mail: alfredo.huete@uts.edu.au

- Seddon, A. W. R., Macias-Fauria, M., Long, P. R., Benz, D. & Willis, K. J. *Nature* **531**, 229–232 (2016).

- Scheffer, M. *et al.* *Nature* **461**, 53–59 (2009).
- Huxman, T. E. *et al.* *Nature* **429**, 651–654 (2004).
- Ponce-Campos, G. E. *et al.* *Nature* **494**, 349–352 (2013).
- Nemani, R. R. *et al.* *Science* **300**, 1560–1563 (2003).
- Guan, K. *et al.* *Nature Geosci.* **8**, 284–289 (2015).
- Ma, X., Huete, A., Moran, S., Ponce-Campos, G. & Eamos, D. *J. Geophys. Res.* **120**, 2036–2052 (2015).
- Musavi, T. *et al.* *Int. J. Appl. Earth Observ. Geoinf.* **43**, 119–131 (2015).
- Levine, N. M. *et al.* *Proc. Natl Acad. Sci. USA* **113**, 793–797 (2016).
- De Keersmaecker, W. *et al.* *Glob. Change Biol.* **20**, 2149–2161 (2014).
- Ogle, K. *et al.* *Ecol. Lett.* **18**, 221–235 (2015).

This article was published online on 17 February 2016.

NEURODEVELOPMENT

Regeneration switch is a gas

Nitric oxide gas has now been found to act as a switch during developmental remodelling of axonal projections from neurons: high gas levels promote the degeneration of unwanted axons and low levels support subsequent regrowth.

TAKESHI AWASAKI & KEI ITO

To create fantastic bonsai trees, a bonsai master prunes unwanted branches and promotes the growth of new ones with careful timing. Similarly, neuronal projections called axons must undergo proper and timely pruning and regrowth in the brain to produce functional neuronal circuits¹. Failure of this process has been associated with autism and schizophrenia^{2,3}. Until now, the way in which neurons transition between degenerative and regenerative states has been mysterious, but, writing in *Cell*, Rabinovich *et al.*⁴ report that the switch is mediated by levels of nitric oxide (NO) gas.

The mushroom body (MB) is a brain region in the fruit fly *Drosophila melanogaster* that is involved in associative learning and memory. During early pupal development, when larvae undergo metamorphosis into flies, the distal branches of MB axons are eliminated and then regrow, adopting different conformations that better serve the adult lifestyle. As such, MB axons offer an excellent model system in which to untangle the mechanisms that underlie neuronal remodelling⁵. Research on this system has provided a good understanding of axon degeneration^{6–8}, but axon regeneration and the mechanisms that control the transition between the two states have not been well studied.

The group that performed the current study previously showed that, in *D. melanogaster*, a nuclear receptor protein called UNF is essential

for axon regrowth⁹. In mice, the equivalent protein forms a dimer with another nuclear receptor, REV-ERB (ref. 10). Rabinovich *et al.* found that the fruit-fly equivalent of REV-ERB, a protein called E75, is also essential for axon regrowth. It has been proposed¹¹ that haem molecules bind to each of UNF and E75, and that haem also binds to NO gas. In addition, NO levels modulate the activity of E75 (ref. 12). The authors therefore investigated whether NO is involved in axon regrowth during MB remodelling.

Using MB neurons in culture, Rabinovich *et al.* reduced NO levels by inhibiting the activity of the enzyme that catalyses NO production, NO synthase (NOS), either chemically or by inhibiting transcription of the *NOS* gene. Both treatments promoted regrowth of MB axons. By testing the physical interaction between UNF and E75, the researchers found evidence that the proteins interacted when NO was depleted, but not under normal conditions. Thus, they suggest that UNF and E75 form dimers that promote axon regrowth, but can do so only when NO levels are low. Moreover, depleting NOS *in vivo* caused not only precocious regrowth but also defective pruning, demonstrating the need for high NO levels during the degenerative phase of remodelling.

Next, the authors showed that NO levels in MB neurons undergo dynamic change during normal remodelling, being high during pruning and low during regrowth. However,

levels of NOS messenger RNA and NOS remained unchanged during the transition between states. How, then, is the level of NO controlled? The NOS DNA sequence generates several mRNA isoforms, and Rabinovich *et al.* found that expression of at least one of these, which encodes a truncated form of NOS, coincided with regrowth but not pruning. NOS proteins must bind together into dimers to act enzymatically, so the production of truncated NOS isoforms might limit the capacity of even full-length NOS proteins to form functional dimers, severely decreasing NO synthesis.

To test this, the authors overexpressed full-length NOS in mutant MB neurons that lacked all NOS isoforms. As predicted, axon regrowth was drastically delayed. By contrast, regrowth was normal when full-length NOS was overexpressed in healthy MB neurons expressing the truncated NOS isoform. Rabinovich *et al.* therefore concluded that expression of the truncated NOS isoform disrupts the formation of functional NOS dimers, causing a rapid drop in NO levels. This change allows the formation of UNF-E75 dimers, which activate downstream signalling pathways to promote axon regrowth (Fig. 1). How expression of the short isoform is controlled over time remains unclear, and identification of the underlying regulatory mechanisms will be the key to deciphering this.

NO is known to regulate the synaptic connections between neurons¹³, changing their strength in a gradual, activity-dependent manner. This regulation primarily involves the classic NO signalling pathway, in which NO induces production of cyclic GMP molecules through activation of the enzyme soluble guanylate cyclase, leading to local changes in synaptic regions of the cell. By contrast, Rabinovich *et al.* describe a process in which NO exerts acute, switch-like regulation. This difference can be explained by the fact that the regrowth switch acts not through the classic pathway, but through UNF and E75 — transcriptional regulators that probably act in the nucleus to modulate the expression of many genes after dimerization. This is a role for NO that was previously unknown.

Pharmacological inhibition of NO-cGMP signalling in photoreceptor neurons of the pupal fly brain induces the formation of disorganized and overextended axons¹⁴. It is intriguing that, even in the pupal brain, NO has different roles in different neurons and acts through different downstream targets. Thus, rapid changes in NO levels might simultaneously activate several developmental programs according to cell type.

How high NO levels promote axon degeneration remains unclear. During metamorphosis, NO-mediated E75 inhibition activates another nuclear receptor, FTZ-F1 (ref. 12). During MB remodelling, FTZ-F1 mutant MB neurons show pruning defects⁷, raising the possibility that NO-mediated E75 inactivation is required

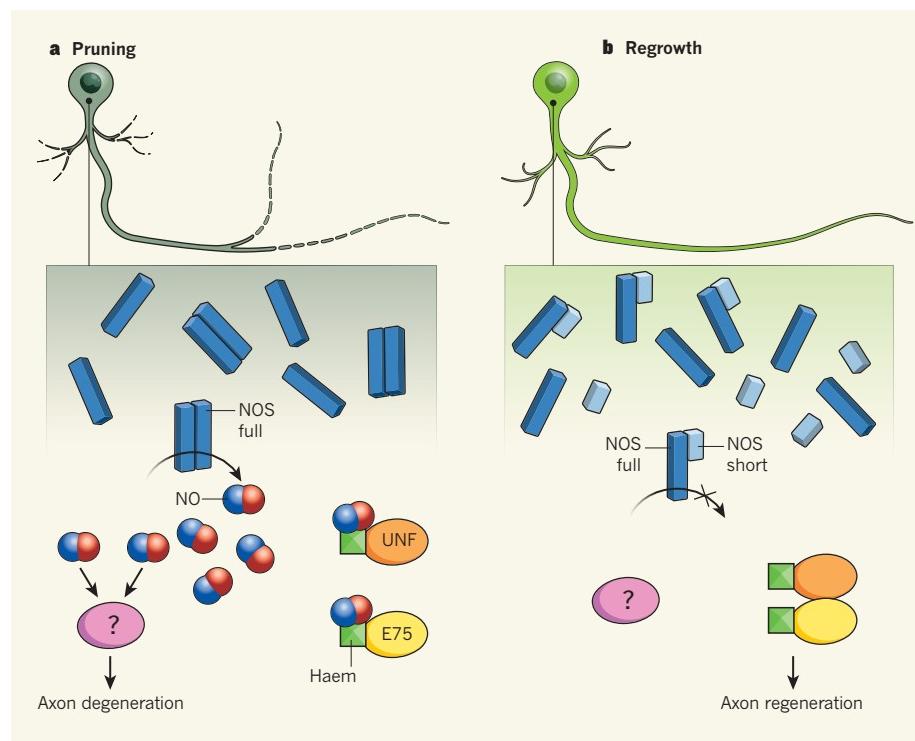


Figure 1 | A switch that controls neuronal remodelling. Neuronal projections called axons in the mushroom body (MB) region of fruit-fly brains undergo pruning and subsequent regrowth during development. Rabinovich *et al.*⁴ report that levels of nitric oxide (NO) gas mediate the switch between degeneration and regeneration. **a**, During pruning, MB neurons produce only the full-length version of the enzyme NO synthase (NOS), which forms dimers that catalyse NO production. The authors propose that NO binds to haem molecules, which are associated with the nuclear proteins UNF or E75. This suppresses the formation of UNF-E75 dimers, which would promote axon regrowth. High NO levels promote axon degeneration through as yet unidentified mechanisms. **b**, During axon regrowth, production of a short NOS isoform results in the formation of dysfunctional dimers, which, in turn, causes a decrease in NO levels. This drop in NO allows UNF and E75 to dimerize and thus activate signalling pathways that trigger axon regeneration.

for pruning. However, Rabinovich *et al.* found that E75 mutants also had modest pruning defects. Thus, in the pruning phase, NO signalling probably acts through a different pathway (Fig. 1). Many signalling molecules have essential roles in MB pruning⁷, including TGF- β , the steroid hormone ecdysone and FTZ-F1. Testing the interactions between these signals and NO at high NO levels would help to reveal how NO promotes axon degeneration.

Given that mammals also have versions of the UNF, E75 and NOS proteins, and that the first two act as a dimer whose formation is probably affected by NO levels, it is plausible that a similar, albeit slightly different, molecular mechanism is found in humans, perhaps functioning during developmental remodelling in the brain. A connection between neurological disorders and defective neurodevelopmental remodelling is now becoming evident¹. As such, it is worth investigating whether the NO switch acts in species beyond fruit flies. ■

Takeshi Awasaki is in the Department of Biology, Kyorin University School of Medicine, Mitaka, Tokyo 181-8611, Japan. Kei Ito is

at the Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi, Tokyo 113-0032, Japan.
e-mails: awasaki@ks.kyorin-u.ac.jp; itokei@iam.u-tokyo.ac.jp

- Riccomagno, M. M. & Kolodkin, A. L. *Annu. Rev. Cell Dev. Biol.* **31**, 779–805 (2015).
- Glausier, J. R. & Lewis, D. A. *Neuroscience* **251**, 90–107 (2013).
- Tang, G. *et al.* *Neuron* **83**, 1131–1143 (2014).
- Rabinovich, D., Yaniv, S. P., Alyagor, I. & Schuldiner, O. *Cell* **164**, 170–182 (2016).
- Lee, T., Lee, A. & Luo, L. *Development* **126**, 4065–4076 (1999).
- Awasaki, T. & Ito, K. *Curr. Biol.* **14**, 668–677 (2004).
- Yu, F. & Schuldiner, O. *Curr. Opin. Neurobiol.* **27**, 192–198 (2014).
- Awasaki, T., Huang, Y., O'Connor, M. B. & Lee, T. *Nature Neurosci.* **14**, 821–823 (2011).
- Yaniv, S. P., Issman-Zecharya, N., Oren-Suissa, M., Podbilewicz, B. & Schuldiner, O. *Curr. Biol.* **22**, 1774–1782 (2012).
- Cheng, H. *et al.* *Hum. Mol. Genet.* **13**, 1563–1575 (2004).
- Reinkind, J. *et al.* *Cell* **122**, 195–207 (2005).
- Cáceres, L. *et al.* *Genes Dev.* **25**, 1476–1485 (2011).
- Hardingham, N., Dachtler, J. & Fox, K. *Front. Cell. Neurosci.* **7**, 190 (2013).
- Gibbs, S. M. & Truman, J. W. *Neuron* **20**, 83–93 (1998).

This article was published online on 2 March 2016.

A hippocampal network for spatial coding during immobility and sleep

Kenneth Kay¹, Marielena Sosa¹, Jason E. Chung¹, Mattias P. Karlsson¹, Margaret C. Larkin¹ & Loren M. Frank^{1,2}

How does an animal know where it is when it stops moving? Hippocampal place cells fire at discrete locations as subjects traverse space, thereby providing an explicit neural code for current location during locomotion. In contrast, during awake immobility, the hippocampus is thought to be dominated by neural firing representing past and possible future experience. The question of whether and how the hippocampus constructs a representation of current location in the absence of locomotion has been unresolved. Here we report that a distinct population of hippocampal neurons, located in the CA2 subregion, signals current location during immobility, and does so in association with a previously unidentified hippocampus-wide network pattern. In addition, signalling of location persists into brief periods of desynchronization prevalent in slow-wave sleep. The hippocampus thus generates a distinct representation of current location during immobility, pointing to mnemonic processing specific to experience occurring in the absence of locomotion.

The hippocampus is essential for memory and spatial navigation, but we still do not know how these cognitive functions are made possible by the hippocampal neural circuit. Examination of hippocampal neural activity during naturalistic behaviours yields a landmark clue: during locomotion, hippocampal principal neurons, known as ‘place’ cells, fire when subjects traverse discrete locations in space^{1,2}. Place cell firing thus provides an internal representation of space understood to be required for both spatial navigation and episodic memory^{1,3,4}. Yet despite extensive study of place cells, it remains an open question whether place firing reliably persists in the absence of movement, and, if so, whether distinct hippocampal neurons and network mechanisms are engaged. This matter is of fundamental importance as immobility punctuates spatial exploration^{5,6} and features in a range of behaviours dependent on the hippocampus^{1,7,8}, including contextual fear conditioning⁹ and trace conditioning¹⁰.

Previous work focusing on hippocampal neural activity during immobility has identified firing related to past and even upcoming experience^{11–15}. Most striking is the observation that place cells during immobility often re-activate in brief bouts at locations outside of their spatial receptive fields. These brief re-activations occur in conjunction with hippocampal sharp wave-ripples (SWRs)^{16,17}, massively synchronous network events lasting ~100 ms and reflecting high firing rates and strong excitatory drive in hippocampal subregions CA1, CA3, and DG^{1,16–20}. Recent work indicates that place cell firing during SWRs frequently represents spatial sequences remote from the animal’s current position^{14,17,21–23}, further raising the question of whether and how the hippocampus sustains a representation of current position during immobility.

A distinct neuron population at CA2

We recorded neural activity in hippocampal subregions CA1, CA2, CA3, and DG (Fig. 1a) in rats engaged in a hippocampus-dependent spatial memory task^{21,24}, with interleaved rest sessions in an enclosed box. In the task, subjects were trained to alternate between each of three locations (reward wells) in a W-shaped maze (Extended Data Fig. 1a). In examining single neuron (unit) activity, we observed principal units (Fig. 1b) that fired at continuously high rates during

immobility (Extended Data Fig. 2a). This basic observation led us to investigate hippocampal activity in this behavioural state.

We first found that, although SWRs were prominent during immobility, SWR periods comprised only a small proportion of time spent immobile (<10%, Extended Data Fig. 2b), suggesting that SWRs could not account for the observed continuous firing. Next, in examining unit firing at the time of SWRs, we were struck by putative principal units recorded in CA2 that consistently decreased firing during both task and rest SWRs, in contrast to CA1 and CA3 principal units, which increased firing (Fig. 1c, d). Indeed virtually all CA1 and CA3 principal units fired more during SWRs (permutation tests at $P < 0.05$, CA1: 478 out of 489 units, CA3: 271 out of 276 units), while a substantial proportion of putative principal units recorded at CA2 sites were either inhibited or showed no change in firing rate during SWRs, despite otherwise firing hundreds to thousands of spikes during single task epochs (84 out of 226 CA2 site units, with 56 of 84 significantly inhibited during SWRs; Fig. 1e and Extended Data Fig. 3). We termed these atypical units at CA2 sites ‘N’ units (non-positively modulated by SWRs) to distinguish them from conventionally responding ‘P’ units (positively modulated).

N units fire more during immobility

We next examined the relationship of N unit firing to ongoing behaviour. We found that N units fired mainly at low movement speeds and during immobility (Fig. 2a). To characterize this relationship, we first evaluated the correlation between unit firing rate and speed (Fig. 2b). The CA1 and CA3 unit populations both showed overall positive correlation, consistent with previous reports^{25–27} (Pearson’s r , firing rate versus log speed; mean ± s.d.; CA1: 0.11 ± 0.10 , CA1 versus 0, $P < 10^{-58}$, signed-rank; CA3: 0.06 ± 0.11 , CA3 versus 0, $P < 10^{-14}$, signed-rank). Remarkably, the CA2 N and CA2 P unit populations showed dramatically different distributions: P units were positively correlated while N units were almost exclusively negatively correlated (mean ± s.d.; CA2 P: 0.10 ± 0.13 , CA2 P versus 0, $P < 10^{-11}$, signed-rank; CA2 N: -0.10 ± 0.09 , CA2 N versus 0, $P < 10^{-10}$, signed-rank; CA2 N versus CA2 P, $P < 10^{-19}$, rank-sum). N units also fired at higher rates than all other unit populations during immobility (Fig. 2c). These findings indicated a fundamental distinction between N units and classic hippocampal place cells.

¹UCSF Center for Integrative Neuroscience and Department of Physiology, University of California San Francisco, California 94158, USA. ²Howard Hughes Medical Institute, University of California San Francisco, California 94158, USA.

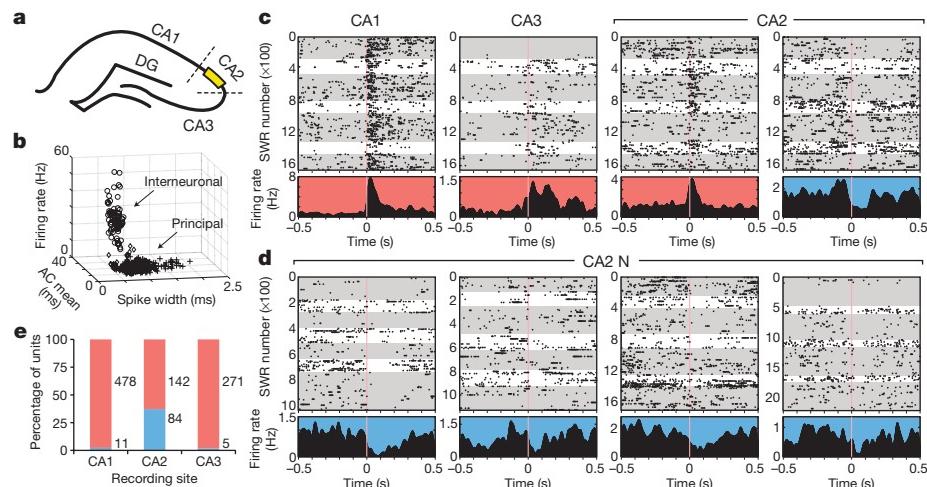


Figure 1 | Distinct hippocampal neuron population at CA2.

a, Diagram of hippocampal recording sites. Recording locations were designated as CA2 sites if found to overlap with the CA2 cytoarchitectural locus⁵⁰. A molecularly defined CA2 region is shown as a yellow band. Additional description is provided in Extended Data Fig. 1.

b, Classification of putative principal versus interneuronal units. Shown is a scatter plot of all hippocampal neural units in the task data set for the three features used to classify units in this study. AC mean: autocorrelation function mean. Open circles: interneuronal ($n = 78$); plus symbols: principal ($n = 991$); open diamonds: unclassified ($n = 21$).

c, Firing aligned to SWRs ($t = 0$: time of SWR onset) in four simultaneously recorded hippocampal putative principal units. Upper sections: SWR-triggered spike rasters (black dots). Grey zones demarcate rest epochs; white zones demarcate task epochs. Lower sections: peri-SWR time histogram (PSTH; 1-ms bins) smoothed with a Gaussian kernel ($\sigma = 10$ ms). Red background indicates increased firing during SWRs; blue background indicates lack of increase. The CA2 site units were recorded on the same tetrode.

d, Firing aligned to SWRs in four example CA2 N units. Each unit was recorded from a different subject.

e, Percentages of P (red) versus N (blue) units at CA1, CA2, and CA3 recording sites. Numbers correspond to unit counts.

N units signal location during immobility

We next assessed whether N units showed spatial firing. We found that N units showed less spatial coverage than the other unit populations (Fig. 3a, b and Extended Data Fig. 4). In contrast, CA2 P units typically showed large spatial fields, consistent with recent reports^{28–30}.

In conjunction with low spatial coverage, N unit firing maps showed concentrated firing at locations where subjects were immobile

to SWRs ($t = 0$: time of SWR onset) in four simultaneously recorded hippocampal putative principal units. Upper sections: SWR-triggered spike rasters (black dots). Grey zones demarcate rest epochs; white zones demarcate task epochs. Lower sections: peri-SWR time histogram (PSTH; 1-ms bins) smoothed with a Gaussian kernel ($\sigma = 10$ ms). Red background indicates increased firing during SWRs; blue background indicates lack of increase. The CA2 site units were recorded on the same tetrode. **d**, Firing aligned to SWRs in four example CA2 N units. Each unit was recorded from a different subject. **e**, Percentages of P (red) versus N (blue) units at CA1, CA2, and CA3 recording sites. Numbers correspond to unit counts.

(Fig. 3a and Extended Data Fig. 4c). To quantify possible spatial specificity in firing during immobility, we focused on firing at the maze reward wells since immobility at these locations was common across all subjects. Our analysis revealed that individual N units characteristically fired at specific single reward wells, while remaining silent at the others (Fig. 3c, d and Extended Data Fig. 5a). The location of N unit firing did not require direct association with reward since spatially specific firing was also observed at other maze locations

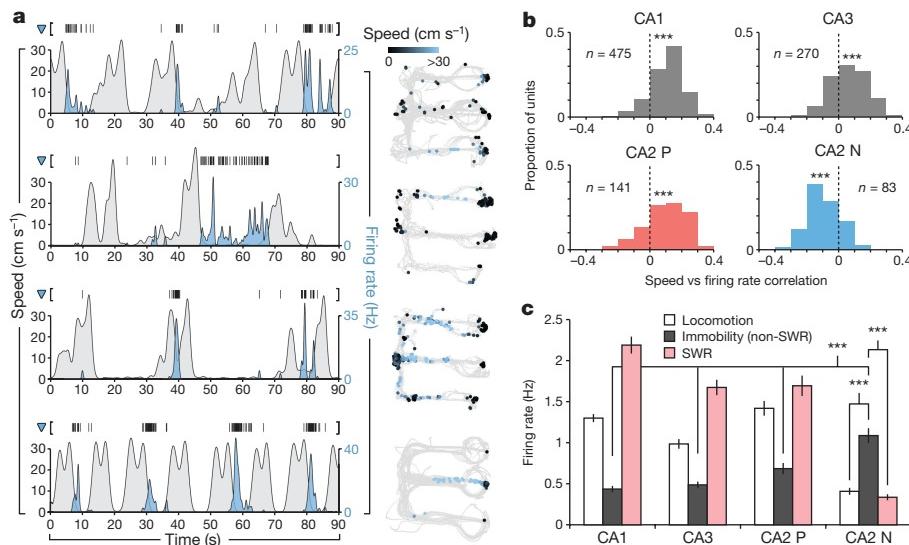


Figure 2 | N units fire more at low speeds and during immobility.

a, Firing of four example CA2 N units during task behaviour. Each row corresponds to an N unit, with spike rasters plotted above the traces. Left y axis and grey fill trace: head speed (cm s^{-1}) of the subject. Right y axis and blue fill trace: instantaneous firing rate (Hz). Right panels: spatial firing maps from corresponding task epochs. Grey: positions visited; coloured points (darker colour values at lower speeds): positions at which firing occurred, with each point opaque and plotted chronologically.

b, Distribution of correlations (Pearson's r) between firing rate and log speed for each hippocampal unit population. *** $P < 0.001$ (versus $r = 0$).

c, Mean firing rates during task epochs (mean \pm s.e.m.; number of units: CA1: 478, CA3: 271, CA2 P: 142, CA2 N: 84). Across unit populations, N units showed the highest firing rates during non-SWR immobility (Kruskal–Wallis ANOVA, Tukey's post hoc tests for CA2 N greater than each other population, $P < 0.001$). Moreover, N unit firing was higher during non-SWR immobility than during locomotion ($P < 10^{-10}$, signed-rank) and also SWRs ($P < 10^{-12}$, signed-rank). *** $P < 0.001$.

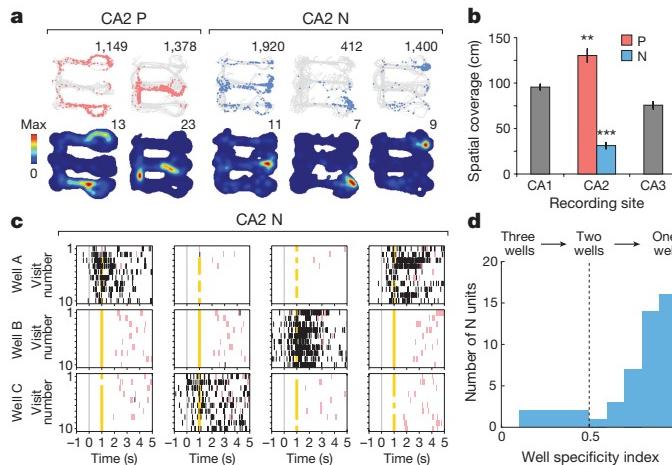


Figure 3 | N units signal current location during immobility. **a**, Spatial firing maps of five example CA2 site units. Each column corresponds to a unit. Upper row: positions visited (grey) and positions where the unit fired (coloured points: P units in red, N units in blue). Total number of spikes is reported at upper right. Lower row: occupancy-normalized firing maps. Peak spatial firing rate is reported at upper right. Subjects stopped locomoting at the ends of the maze arms to receive reward and also stopped intermittently elsewhere in the maze (Extended Data Fig. 1a). **b**, Spatial coverage in the hippocampal unit populations (mean \pm s.e.m.; number of units: CA1: 476, CA2 P: 142, CA2 N: 79, CA3: 271). The CA2 N and P unit populations showed the lowest and highest spatial coverages, respectively (Kruskal–Wallis ANOVA, Tukey’s post hoc tests, CA2 P greater than each other population, $P = 0.0015$; CA2 N less than each other population, $P < 10^{-6}$). ** $P < 0.01$; *** $P < 0.001$. **c**, Reward well firing of four example CA2 N units. Each column corresponds to a unit. For each well, the last ten visits (in a task recording epoch) are shown. Grey line: time of well entry ($t = 0$); yellow line: time of reward delivery (omitted in error trials). SWR periods are shown as pink zones. The two leftmost units were recorded simultaneously and on the same tetrode. **d**, Well specificity distribution in the N unit population. Mean \pm s.e.m.: 0.78 ± 0.03 ($n = 53$ units).

(Extended Data Fig. 5b–d; seen previously in Fig. 2a and Extended Data Fig. 4c). These findings indicate that N unit firing constitutes a precise neural code for location during immobility.

A signature of spatial coding during immobility

We were struck by the fact that the firing pattern of N units not only was unorthodox (Fig. 1), but also had unambiguous behavioural (Fig. 2) and representational (Fig. 3) correlates. We hypothesized that this distinctive firing was the result of an unidentified input pattern in the hippocampus. To evaluate this possibility, we calculated CA2 site (N and P) unit spike-triggered averages (STAs) of hippocampal local field potential (LFP)¹⁸, analysing locomotor and immobility periods separately (Fig. 4a).

In contrast to STAs from locomotor periods (characterized by the expected ~8 Hz theta frequency modulation^{18,31}, Extended Data Fig. 6), STAs from non-SWR immobility periods (Fig. 4b, c and Extended Data Fig. 7a) showed that N units fired at the time of a positive transient LFP pattern lasting ~200 ms. The pattern was smallest on the parent electrode in CA2, larger in CA3, and largest at DG, suggesting broad engagement of the hippocampal circuit. Furthermore, unlike N units, P units showed a mean STA characterized by a negative transient similar to the canonical sharp wave transient of SWRs³² (Fig. 4b, c).

Power spectral analysis (Fig. 4d) further specified the contrasting LFP patterns. The power spectral density (PSD) of CA2 N and P unit immobility STAs and of SWR sharp waves showed fundamental frequencies <5 Hz, a bandwidth distinct to that of theta^{18,31} (5–11 Hz). In agreement, STAs of LFP filtered at 1–4 Hz showed the same pattern of transients as in the wide-band STAs (Extended Data Fig. 7a).

indicating that filtering at 1–4 Hz effectively isolates the large-amplitude transients associated with CA2 N units, CA2 P units, and SWRs. The N unit STA pattern exceeded 0 mV (Extended Data Fig. 7b, c), in fundamental contrast to SWR sharp waves³², which manifested as negative transients. Thus, N units fired in association with an LFP pattern distinct from canonical hippocampal LFP patterns^{1,17,18} (theta and SWRs). We termed this pattern ‘N wave’ (N unit-identified wave), a ~200 ms LFP transient with positive polarity at hippocampal recording sites (specifically CA2, CA3, and DG principal cell layers) at which SWR sharp waves are negative.

We then asked whether neurons outside of CA2 were also N wave-coupled. We identified N wave-coupled units in CA1, CA3, and DG (Fig. 4e–i and Extended Data Figs 7d–g, 8, 9), indicating that the N wave reflects a hippocampus-wide network pattern. Critically, a distinct subset of principal units was N-wave coupled (CA1: 50 units, CA3: 34 units, Fig. 4g–i and Extended Data Figs 8 and 9). As with CA2 N units, these units fired more during immobility than during movement (Extended Data Fig. 8b) and showed unequivocal location-specific firing during immobility (Fig. 4g, i and Extended Data Figs 8d, e and 9), thereby linking the N wave network pattern to spatial coding during immobility across the hippocampus.

Hippocampal spatial coding in sleep

Does spatial coding during immobility also occur under quiescent behavioural conditions? Intriguingly, past work has shown that, during slow-wave sleep, ~5% of CA1 place cells continuously fire during episodes in which hippocampal neural activity becomes highly desynchronized, reflected by low-amplitude LFP³³. In this sleep state, termed small-irregular activity (SIA)^{1,33,34}, CA1 place cells were found to signal the location where the subject fell asleep (nesting position)³³. Recent findings show that CA2 neurons send strong excitatory input to CA1 (refs 35–37), raising the possibility that coding of nesting position is staged upstream in CA2.

To test this possibility, we evaluated hippocampal neural activity during rest sessions. First, during sleep, we observed periods of high-amplitude LFP, corresponding to a hippocampal sleep state dominated by SWRs (termed LIA)^{1,18,33,34}, frequently interrupted by periods of low-amplitude LFP in which the subject did not rouse, which we identified as periods of SIA (Fig. 5a). Next, in examining unit firing during sleep, we observed striking instances in which N units fired preferentially during SIA periods, falling silent during LIA (Fig. 5b). Analogously to awake immobility in the task (Fig. 2c), the N unit population fired at higher rates than all other unit populations during SIA (green, Fig. 5c) and also during awake immobility in the rest environment (dark grey, Fig. 5c). However, unlike the task condition, there was no significant overall correlation between firing rate and speed for N units during awake periods in the rest environment (Extended Data Fig. 10a), indicating that properties of the task maze or the cognitive demands of the task have essential roles in regulating N unit firing.

We then asked whether N units represented locations in the rest environment. We found that N units showed spatially specific firing during awake periods (Fig. 5d, Extended Data Fig. 10b) that persisted in awake immobility periods (Extended Data Fig. 10c–i) and furthermore into SIA: specifically, the CA1 and N unit populations met dual criteria for nesting position coding during SIA, while the CA3 population unexpectedly failed both criteria (criteria in Supplementary Methods; Fig. 5e, f and Extended Data Fig. 10j–l). In addition, during awake immobility in the rest environment, the N unit population showed a dominant coupling to the N wave network pattern, suggesting similar or equivalent circuit mechanisms underlying spatial firing during immobility in quiescent conditions as spatial firing during immobility in the task (Extended Data Fig. 10m).

Discussion

These findings identify a distinct hippocampal network at the anatomical (Fig. 1), behavioural (Fig. 2), representational (Fig. 3), and neural

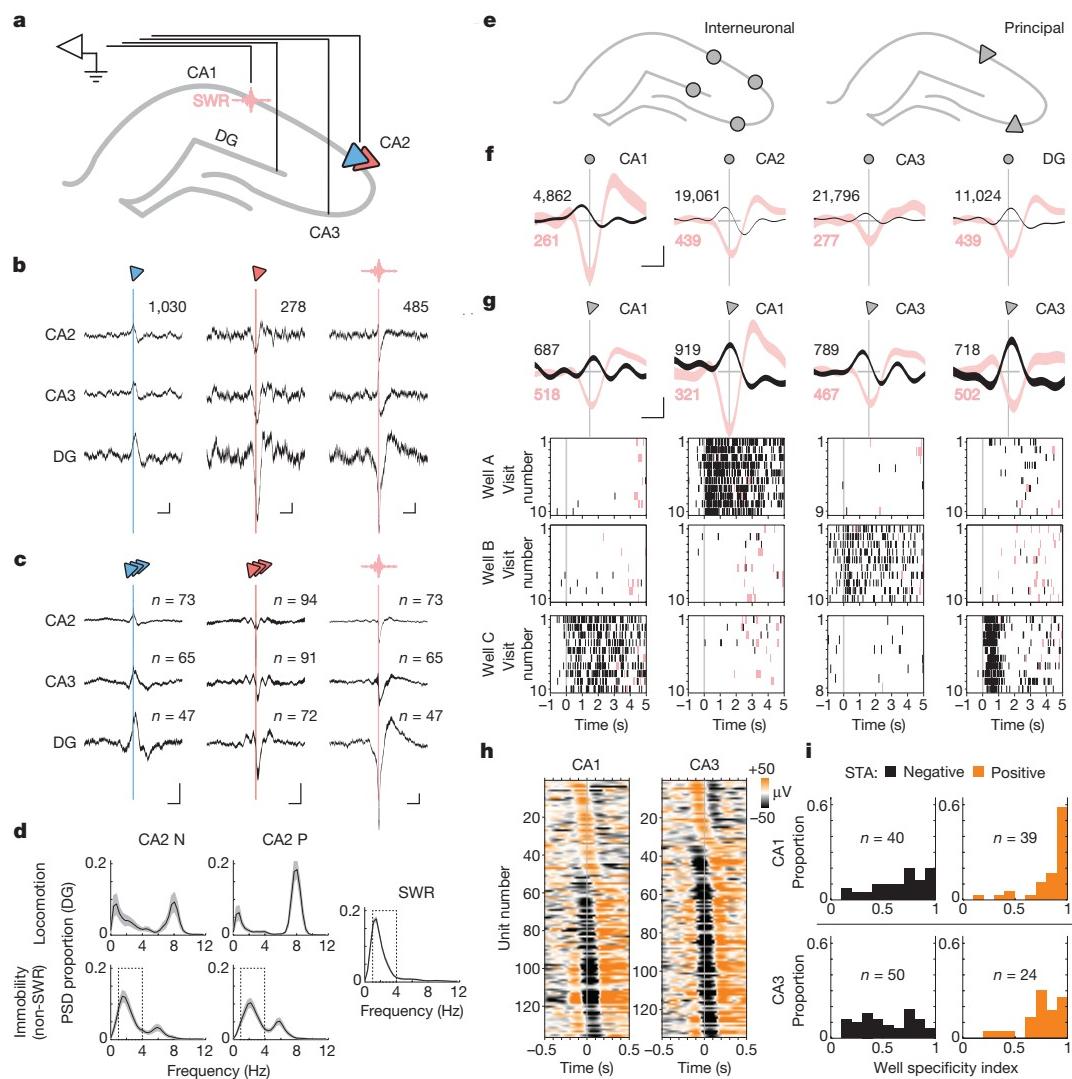


Figure 4 | A novel hippocampal network pattern marks spatial coding during immobility. **a**, Schematic of recording configuration. SWRs (pink symbol) were detected with CA1 site electrodes, while wide-band LFP was taken from CA2, CA3, and DG site electrodes. Blue and red symbols refer to CA2 N and CA2 P units, respectively, analysed in **b–d**. **b**, Example CA2 N (blue symbol, first column) and CA2 P (red symbol, second column) unit spike-triggered average (STA) and SWR-triggered average (RTA; pink symbol, third column) of hippocampal CA2, CA3, and DG LFP from non-SWR immobility periods. Vertical lines indicate the time of spiking (STA) or time of SWR (RTA). The two units were recorded simultaneously and on the same tetrode. SWRs averaged in the RTA were detected in the same recording epochs as the units. The total number of events averaged is reported at upper right. Trace width indicates \pm s.e.m. over single LFP traces. Trace length: 2 s. Scale bars: x , 250 ms; y , 100 μ V. **c**, Mean STAs for CA2 N and CA2 P unit populations for non-SWR immobility periods. The mean RTA was calculated from single RTAs matching the recording epochs of N unit STAs, and thus have the same sample size. Trace width indicates \pm s.e.m. over unit STAs/RTAs. Trace length: 2 s. Scale bars: x , 250 ms; y , 100 μ V. **d**, Power spectral density (PSD) of STAs and RTAs of DG LFP. The mean PSD is plotted as a black line, with \pm s.e.m. over single averages plotted in grey (locomotor periods: CA2 N, $n = 39$ units; CA2 P, $n = 85$; non-SWR immobility periods: CA2 N, $n = 47$; CA2 P, $n = 72$; RTAs matched to CA2 N units: $n = 47$). **e**, Schematic of additional hippocampal neurons analysed with STAs. Interneuronal units (left, grey circles) recorded in the principal cell layers of CA1, CA2, CA3 and DG analysed in **f** and Extended Data Fig. 7d–g. Principal units (right, grey triangles) recorded in CA1 and CA3 analysed in **g–i** and Extended Data Figs 8 and 9. STAs in **f–i**

were taken for 1–4 Hz LFP, analysing spikes from non-SWR immobility periods. **f**, N wave firing in four example interneuronal units. Plotted are unit STAs and RTAs. Trace width indicates \pm s.e.m. (STA, black) or \pm 2 s.e.m. (RTA, pink) over single LFP traces. Vertical line: time of spiking (STAs) or SWRs (RTAs). The hippocampal subregion in which the unit was recorded is reported at upper right. The number of spikes or SWRs averaged is indicated at upper and lower left, respectively. Trace length: 1 s. Horizontal line (200 ms in length): 0 μ V. DG LFP was used in each example except for the CA3 unit, which used CA3 LFP. Scale bars: x , 200 ms; y , 50 μ V for STA, 100 μ V for RTA. **g**, N wave firing and well specificity in four example CA1/CA3 principal units. Top: unit STAs and RTAs, following the plotting conventions in **f**. DG LFP was used in each example. Bottom: well firing rasters correspond to each unit. Grey line: time of well entry ($t = 0$). SWR periods plotted as pink zones. **h**, CA1 and CA3 unit STAs. Colour indicates voltage. For each unit, LFP (1–4 Hz) from DG, CA3, or CA2 (in decreasing order of preference) was used. Unit STAs were grouped by polarity at the time of spiking ($t = 0$) and sorted by the time of the local extremum (peak for positive; trough for negative) nearest the time of spiking. Units with positive voltage peaks at the time of spiking were classified as N wave-coupled. **i**, Well specificity distributions for CA1 and CA3 principal unit populations classified by STA. For both CA1 and CA3 populations, units with positive STAs (N wave-coupled) showed higher well specificity than units with negative STAs (mean \pm s.e.m.; CA1 positive: 0.85 ± 0.03 ; CA1 negative: 0.65 ± 0.04 , CA1 positive versus CA1 negative, $P < 10^{-5}$, rank-sum; CA3 positive: 0.77 ± 0.04 ; CA3 negative: 0.53 ± 0.04 , CA3 positive versus CA3 negative, $P < 0.001$, rank-sum).

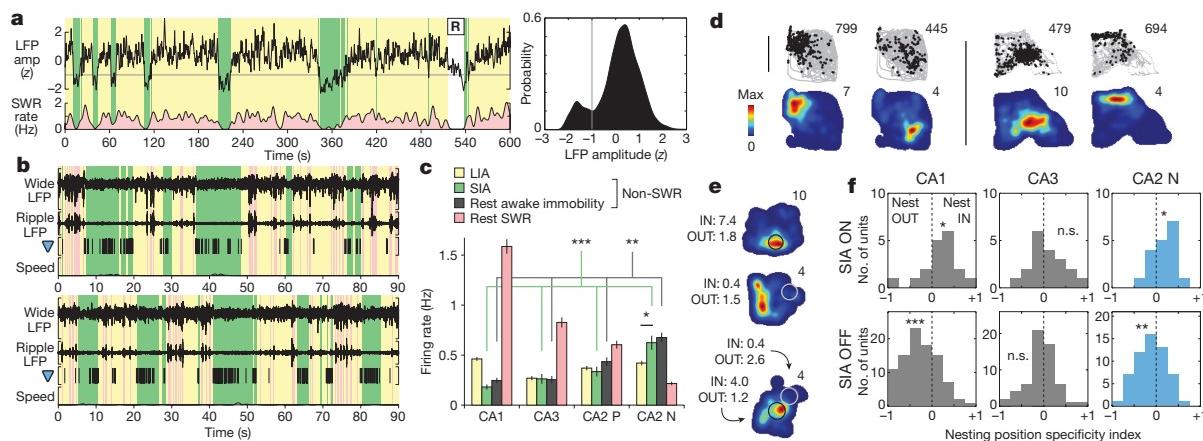


Figure 5 | Hippocampal spatial coding in desynchronized sleep.

a, Detection of sleep states using hippocampal LFP. Left, 10-min trace of aggregate hippocampal LFP amplitude during sleep, with times classified as LIA (yellow), SIA (green), or REM (R) periods. SWR rate was estimated by counting SWRs in 1-s bins and smoothing with a Gaussian ($\sigma = 2$ s). Right, kernel density estimate (Gaussian kernel, $\sigma = 0.1$) of aggregate hippocampal LFP amplitude during non-REM sleep for the recording epoch from which the plotted trace was taken. Grey line: amplitude threshold used to distinguish SIA (below threshold) and LIA (above threshold) periods. **b**, Sleep firing in two example CA2 N units. Top traces: wide-band LFP (Wide, 0.5–400 Hz, scale bar: 2 mV) and ripple-band LFP (Ripple, 150–250 Hz, scale bar: 300 μ V) traces from a simultaneous recording in CA1. SWR, LIA, and SIA periods are plotted as pink, yellow, and green zones, respectively. Grey-filled trace (y axis: 0 to 10 cm s^{-1}): head speed. Subsequent analysis in **d–f** indicated that SIA firing was dependent on whether the location at which the animal slept was near the spatial firing field of the CA2 N unit. **c**, Mean firing rates during rest epochs (mean \pm s.e.m.; number of units: CA1: 400, CA3: 220, CA2 P: 126 units, CA2 N: 76 units). CA2 N units fired more during SIA than LIA ($P = 0.011$, signed-rank) and at higher rates than other unit populations during SIA periods (green) and during awake immobility periods (grey) (Kruskal–Wallis ANOVA, Tukey's post hoc tests; $P < 0.001$ for SIA; $P = 0.0051$ for awake immobility). As in Fig. 2c, these comparisons indicate population-level engagement in sleep states, encompassing both higher and lower rate firing as a result of spatially

specific firing in single units. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

d, Example spatial firing maps of two pairs of simultaneously recorded CA2 N units in the rest environment. Data from waking periods plotted. Upper plots: positions visited (grey) and positions where the unit fired (black points). Total number of spikes is reported at upper right. Lower plots: occupancy-normalized firing maps. Peak spatial firing rate is reported at upper right. Scale bar: 20 cm. **e**, Three example CA2 N units coding for nesting position. Shown are occupancy-normalized firing maps from awake periods in a rest recording epoch. Indicated on each map is the nesting position (circle, 5 cm radius) of the subject for a sleep period detected in the same recording epoch. For a given sleep period, the unit was classified either as SIA ON (>2 Hz firing rate during SIA; black circle) or SIA OFF (<2 Hz; white circle). Reported at left are the mean awake firing rates within (Nest IN) and outside (Nest OUT) the encircled nesting region. In the third example, two distinct nesting positions corresponding to two distinct sleep periods were observed. **f**, Nesting position specificity index distribution in CA1, CA3, and CA2 N unit populations. The CA1 and CA2 N populations met dual criteria (see Methods) for nesting position coding, while the CA3 unit population did not. Mean \pm s.e.m.: CA1, SIA ON ($n = 18$ units): 0.18 ± 0.09 , $P = 0.043$; CA1, SIA OFF ($n = 92$): -0.26 ± 0.04 , $P < 10^{-6}$; CA3, SIA ON ($n = 19$): 0.09 ± 0.09 , $P = 0.47$; CA3, SIA OFF ($n = 58$): -0.04 ± 0.04 , $P = 0.50$; CA2 N, SIA ON ($n = 18$): 0.18 ± 0.06 , $P = 0.020$; CA2 N, SIA OFF ($n = 57$): -0.12 ± 0.04 , $P = 0.0087$. All statistical tests were signed-rank. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; n.s., not significant at $P < 0.05$.

circuit (Figs 1 and 4) levels, and also indicate its activation in sleep (Fig. 5). In the awake animal, neural firing in this network is marked by a distinct hippocampal network pattern (N wave), occurs during immobility in subsets of neurons in CA1, CA2, CA3, and DG, and is location-specific, constituting an explicit neural code for current position. Thus the classic locomotor hippocampal place code switches, during immobility, to an alternative hippocampal neural code that nonetheless maintains spatial specificity.

Past observations of a lack of place cell firing in restrained animals have led to the suggestion that place firing is driven by input correspondent with an animal's preparedness to make limb movements that would displace the animal from its current position, a condition termed 'motor set'^{38,39}. Moreover, in rodents, hippocampal theta has been proposed to be a marker of motor set^{1,34,40}, and thus by inference a marker of hippocampal place firing. Here we observe spatial firing dependent on neither theta nor motor set, indicating that distinct mechanisms can generate spatial firing and in fact do so complementarily.

A neural code for location during awake immobility enables the brain to provide a spatial context to events occurring during immobility such as consumption of food, sensory stimuli, and deliberation, allowing for the formation of location-specific memories when the animal is still. Moreover, we suggest that the various hippocampus-dependent behaviours characterized by immobility^{1,5–10} engage this network, and that activity in this network may correspond to activity seen in human⁴¹, monkey⁴², and bat⁴³ hippocampus, where the theta network pattern occurs less frequently. Importantly, analysis of firing during immobility

has not been prominent in traditional approaches to hippocampal spatial coding, in which behavioural paradigms eliciting continuous locomotion or post hoc exclusion of immobility periods is the norm.

Remarkably, a distinct population of hippocampal neurons located at CA2 (N units) signalled location during not only awake immobility but also sleep. An internal representation of current location active during sleep could adaptively influence representations reactivated in sleep in support of memory consolidation^{44–46}, and, concurrently, could serve to maintain a sleeping animal's bearings despite diminished receptivity to sensory stimuli.

Finally, the localization of N units at CA2 suggests that N units correspond to CA2 neurons, while CA2 P units correspond to intermingling CA1 and CA3 neurons at the CA2 anatomical locus. In parallel with the unique firing pattern of N units, CA2 neurons exhibit a variety of properties unique among hippocampal neurons, including a unique synaptic configuration^{35–37,47,48}. Moreover, a recent study reports suppressed firing in three identified CA2 neurons during SWRs⁴⁹, indicating that N units and CA2 neurons are overlapping populations or in fact identical. Recent work also links CA2 neurons to the generation of time-dependent spatial representations²⁸, spatial pattern completion^{29,30}, and social memory³⁷. These cognitive functions and possibly others may rely on the alternative forms of hippocampal neural activity identified here.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 June 2015; accepted 13 January 2016.

Published online 2 March 2016.

1. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
2. Wilson, M. A. & McNaughton, B. L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993).
3. Buzsáki, G. & Moser, E. I. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neurosci.* **16**, 130–138 (2013).
4. Eichenbaum, H. & Cohen, N. J. Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron* **83**, 764–770 (2014).
5. Eilam, D. & Golani, I. Home base behavior of rats (*Rattus norvegicus*) exploring a novel environment. *Behav. Brain Res.* **34**, 199–211 (1989).
6. Wallace, D. G., Hamilton, D. A. & Whishaw, I. Q. Movement characteristics support a role for dead reckoning in organizing exploratory behavior. *Anim. Cogn.* **9**, 219–228 (2006).
7. Bannerman, D. M. et al. Regional dissociations within the hippocampus—memory and anxiety. *Neurosci. Biobehav. Rev.* **28**, 273–283 (2004).
8. Pentkowski, N. S., Blanchard, D. C., Lever, C., Litvin, Y. & Blanchard, R. J. Effects of lesions to the dorsal and ventral hippocampus on defensive behaviors in rats. *Eur. J. Neurosci.* **23**, 2185–2196 (2006).
9. Maren, S., Phan, K. L. & Liberzon, I. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nature Rev. Neurosci.* **14**, 417–428 (2013).
10. Christian, K. M. & Thompson, R. F. Neural substrates of eyeblink conditioning: acquisition and retention. *Learn. Mem.* **10**, 427–455 (2003).
11. Takahashi, M., Lauwereyns, J., Sakurai, Y. & Tsukada, M. A code for spatial alternation during fixation in rat hippocampal CA1 neurons. *J. Neurophysiol.* **102**, 556–567 (2009).
12. MacDonald, C. J., Carrow, S., Place, R. & Eichenbaum, H. Distinct hippocampal time cell sequences represent odor memories in immobilized rats. *J. Neurosci.* **33**, 14607–14616 (2013).
13. Hattori, S., Chen, L., Weiss, C. & Disterhoft, J. F. Robust hippocampal responsibility during retrieval of consolidated associative memory. *Hippocampus* **25**, 655–669 (2015).
14. Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neurosci.* **14**, 147–153 (2011).
15. Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79 (2013).
16. Buzsáki, G., Horvath, Z., Urioste, R., Hetke, J. & Wise, K. High-frequency network oscillation in the hippocampus. *Science* **256**, 1025–1027 (1992).
17. Buzsáki, G. Hippocampal sharp wave-ripple: a cognitive biomarker for episodic memory and planning. *Hippocampus* **10**, 1073–1188 (2015).
18. Buzsáki, G., Leung, L. W. & Vanderwolf, C. H. Cellular bases of hippocampal EEG in the behaving rat. *Brain Res.* **6**, 139–171 (1983).
19. Csicsvari, J., Hirase, H., Mamiya, A. & Buzsaki, G. Ensemble patterns of hippocampal CA3–CA1 neurons during sharp wave-associated population events. *Neuron* **28**, 585–594 (2000).
20. Penittonen, M., Kamondi, A., Sik, A., Aszady, L. & Buzsaki, G. Feed-forward and feed-back activation of the dentate gyrus *in vivo* during dentate spikes and sharp wave bursts. *Hippocampus* **7**, 437–450 (1997).
21. Karlsson, M. P. & Frank, L. M. Awake replay of remote experiences in the hippocampus. *Nature Neurosci.* **12**, 913–918 (2009).
22. Davidson, T. J., Kloosterman, F. & Wilson, M. A. Hippocampal replay of extended experience. *Neuron* **63**, 497–507 (2009).
23. Gupta, A. S., van der Meer, M. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705 (2010).
24. Kim, S. M. & Frank, L. M. Hippocampal lesions impair rapid learning of a continuous spatial alternation task. *PLoS ONE* **4**, e594 (2009).
25. McNaughton, B. L., Barnes, C. A. & O'Keefe, J. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp. Brain Res.* **52**, 41–49 (1983).
26. Huxter, J., Burgess, N. & O'Keefe, J. Independent rate and temporal coding in hippocampal pyramidal cells. *Nature* **425**, 828–832 (2003).
27. Zheng, C., Bieri, K. W., Trettel, S. G. & Colgin, L. L. The relationship between gamma frequency and running speed differs for slow and fast gamma rhythms in freely behaving rats. *Hippocampus* **25**, 924–938 (2015).
28. Mankin, E. A., Diehl, G. W., Sparks, F. T., Leutgeb, S. & Leutgeb, J. K. Hippocampal CA2 activity patterns change over time to a larger extent than between spatial contexts. *Neuron* **85**, 190–201 (2015).
29. Lu, L., Igarashi, K. M., Witter, M. P., Moser, E. I. & Moser, M. B. Topography of place maps along the CA3-to-CA2 axis of the hippocampus. *Neuron* **87**, 1078–1092 (2015).
30. Lee, H., Wang, C., Deshmukh, S. S. & Knierim, J. J. Neural population evidence of functional heterogeneity along the CA3 transverse axis: pattern completion versus pattern separation. *Neuron* **87**, 1093–1105 (2015).
31. Mizuseki, K., Sirota, A., Pastalkova, E. & Buzsaki, G. Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuro* **64**, 267–280 (2009).
32. Buzsaki, G. Hippocampal sharp waves: their origin and significance. *Brain Res.* **398**, 242–252 (1986).
33. Jarosiewicz, B., McNaughton, B. L. & Skaggs, W. E. Hippocampal population activity during the small-amplitude irregular activity state in the rat. *J. Neurosci.* **22**, 1373–1384 (2002).
34. Vanderwolf, C. H. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr. Clin. Neurophysiol.* **26**, 407–418 (1969).
35. Chevaleyre, V. & Siegelbaum, S. A. Strong CA2 pyramidal neuron synapses define a powerful disinaptic cortico-hippocampal loop. *Neuron* **66**, 560–572 (2010).
36. Kohara, K. et al. Cell type-specific genetic and optogenetic tools reveal hippocampal CA2 circuits. *Nature Neurosci.* **17**, 269–279 (2014).
37. Hitti, F. L. & Siegelbaum, S. A. The hippocampal CA2 region is essential for social memory. *Nature* **508**, 88–92 (2014).
38. Foster, T. C., Castro, C. A. & McNaughton, B. L. Spatial selectivity of rat hippocampal neurons: dependence on preparedness for movement. *Science* **244**, 1580–1582 (1989).
39. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M. B. Path integration and the neural basis of the 'cognitive map'. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
40. Buzsaki, G. Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus* **15**, 827–840 (2005).
41. Watrous, A. J. et al. A comparative study of human and rat hippocampal low-frequency oscillations during spatial navigation. *Hippocampus* **23**, 656–661 (2013).
42. Jutras, M. J., Fries, P. & Buffalo, E. A. Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. *Proc. Natl. Acad. Sci. USA* **110**, 13144–13149 (2013).
43. Ulanovsky, N. & Moss, C. F. Hippocampal cellular and network activity in freely moving echolocating bats. *Nature Neurosci.* **10**, 224–233 (2007).
44. Lee, A. K. & Wilson, M. A. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002).
45. Girardeau, G., Benchenane, K., Wiener, S. I., Buzsaki, G. & Zugaro, M. B. Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neurosci.* **12**, 1222–1223 (2009).
46. Ego-Stengel, V. & Wilson, M. A. Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus* **20**, 1–10 (2010).
47. Dudek, S. M., Alexander, G. M. & Farris, S. Rediscovering area CA2: unique properties and functions. *Nature Rev. Neurosci.* **17**, 89–102 (2016).
48. Rowland, D. C. et al. Transgenically targeted rabies virus demonstrates a major monosynaptic projection from hippocampal area CA2 to medial entorhinal layer II neurons. *J. Neurosci.* **33**, 14889–14898 (2013).
49. Valero, M. et al. Determinants of different deep and superficial CA1 pyramidal cell dynamics during sharp-wave ripples. *Nature Neurosci.* **18**, 1281–1290 (2015).
50. Lorente de Nó, R. Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *Journal für Psychologie und Neurologie* **46**, 113–177 (1934).

Acknowledgements We thank G. Rothschild, D. Liu, J. Yu, S. Jadhav, E. Anderson, P. Sabes, C. Schreiner, M. Stryker, R. Knight, J. O'Doherty, E. Phillips, K. Kay, and B. Mensh for discussion and suggestions, and I. Grossrubatscher, C. Lykken, and S. Harris for technical assistance. This work was supported by the Howard Hughes Medical Institute, an NIH grant (R01 MH090188) and a McKnight Foundation Cognitive and Memory Disorders Award (L.M.F.). K.K. is supported by a Ruth L. Kirchstein National Research Service Award Fellowship (NIH/NIMH) and the UCSF Medical Scientist Training Program.

Author Contributions K.K. and L.M.F. conceived the study. K.K., M.S., J.E.C., M.P.K., and M.C.L., conducted the experiments. K.K. analysed the data. K.K. and L.M.F. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.K. (kenneth.kay@ucsf.edu) or L.M.F. (loren@phy.ucsf.edu).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Subjects, neural recordings, and behavioural task. Eight male Long-Evans rats that were 4 to 9 months old (500–600 g) were food deprived to 85% of their baseline weight and pre-trained to run on a 1-m linear track for liquid reward (sweetened evaporated milk). After subjects alternated reliably, they were implanted with microdrives containing 14 (two subjects), 21 (three subjects), or 30 (three subjects) independently movable four-wire electrodes (tetrodes^{2,51}) targeting dorsal hippocampus (all subjects) and medial entorhinal cortex (one subject), in accordance with University of California San Francisco Institutional Animal Care and Use Committee and US National Institutes of Health guidelines. The minimum number of subjects was established beforehand as four or more, as this is considered to be the minimum necessary to yield data with sufficient statistical power to evaluate the type of effects investigated in this study.

In two subjects, right and left dorsal hippocampus were targeted at AP: −3.7 mm, ML: ± 3.7 mm. In one subject, dorsal hippocampus was targeted at AP: −3.6 mm, ML: +2.2 mm, in addition to medial entorhinal cortex at AP: −9.1, ML: +5.6, at a 10 degree angle in the sagittal plane. Data from these several subjects have been reported in earlier studies^{21,63}. In five subjects, right dorsal hippocampus was targeted at AP: −3.3 to −4.0 mm, ML: +3.5 to +3.9 mm, moreover, in two of these subjects, the septal pole of right hippocampus was targeted with an additional six tetrodes targeted to AP: −2.3 mm, ML: +1.1 mm. Targeting locations were used to position stainless steel cannulae containing 6, 14, 15, or 21 independently driveable tetrodes. The cannulae were circular except in four cases targeting dorsal hippocampus in which they were elongated into ovals (major axis ~2.5 mm, minor axis ~1.5 mm; two subjects with major axis 45° relative to midline, along the transverse axis of dorsal hippocampus; two subjects with major axis 135° relative to midline, along the longitudinal axis of dorsal hippocampus). Data exclusively from tetrodes targeting right dorsal hippocampus were analysed in this study.

In five subjects, viral vectors with optogenetic transgenes were targeted to either right dorsal CA2 (three subjects, AAV2/5-CaMKII-hChR2(H134R)-EYFP, UNC Vector Core, 135 nl at AP: −3.6 mm, ML: +4.2 mm, DV: −4.5 mm), dorsal DG (one subject, AAV2/5-I12B-ChR2-GFP (see ref. 52 for details about the I12B promoter)), 225 nl at AP: −3.75 mm, ML: +2.2 mm, DV: 3.9 mm and AP: −3.75 mm, ML: +1.8 mm, DV: −4.5 mm), or right supramammillary nucleus (one subject, AAV2/5-hSyn-ChETA-EYFP, Penn Vector Core, 135 nl at AP: −4.3 mm, ML: +1.8 mm, and −8.9 mm along a trajectory angled at 6° in the coronal plane). Viruses were delivered during the implant surgery using a glass micropipette (tip manually cut to ~25 μm diameter) attached to an injector (Nanoject, Drummond Scientific). In addition, a driveable optical fibre (62.5/125 μm core/cladding) was integrated in the tetrode microdrive assembly to enable light delivery to hippocampus. This fibre was advanced to its final depth (2.5–3 mm) within 7 days of implantation. Data reported in this study were collected before light stimulation. No overt differences in neural activity were observed in subjects that received virus. In particular, CA2 recording sites reporting heterogeneous unit populations (Extended Data Fig. 3c) were found in subjects either receiving or free of viral vectors.

Over the course of two weeks following implantation, the tetrodes were advanced to the principal cell layers of CA1 (all subjects), CA2 (5 subjects), CA3 (all subjects), and DG (3 subjects). For DG, tetrodes were advanced to the cell layer using a previously described protocol in which the tetrodes were slowly advanced within DG (~10 μm increments) and unit activity monitored over long periods of rest⁵³. DG cell layer was identified by the presence of highly sparsely firing putative principal units. In several subjects, tetrodes were also left in cortex overlying dorsal hippocampus. Neural signals were recorded relative to a reference (REF) tetrode positioned in corpus callosum above right dorsal hippocampus. The REF tetrode reported voltage relative to a ground screw installed in skull overlying cerebellum, and local field potential (LFP) from this tetrode was also recorded. All tetrode final locations were histologically verified (see below).

After 5–7 days of recovery after surgery, subjects were once again food deprived to 85% of their baseline weight, and again pre-trained to run on a linear track for liquid reward. At ~14 days after surgery, six subjects were then introduced to one task W-maze (Extended Data Fig. 1a) and recorded for 3 to 6 days before being introduced to a second task W-maze, located in a separate part of the recording room and rotated 90° relative to the first. On recording days in which the second task W-maze was used, recordings were also conducted in the first task W-maze. In two subjects, recordings were conducted in both task W-mazes on every recording day. The W-mazes were 76 × 76 cm with 7-cm-wide track sections. The two task W-mazes were separated by an opaque barrier.

In each W-maze, subjects were rewarded for performing a hippocampus-dependent continuous alternation task^{21,24,65,70} (Extended Data Fig. 1a). Liquid reward (sweetened evaporated milk) was dispensed via plastic tubing connected to a hole at the bottom of each of the three reward wells (wells A, B, and C), miniature bowls 3 cm in diameter. In three subjects, reward was dispensed via syringes operated manually by an experimenter who was located in a separate part of the recording room. In five subjects, entry of the subject's head into reward wells was sensed by an infrared beam break circuit attached to the well, and reward was automatically delivered by syringe pumps (OEM syringe pumps, Braintree Scientific) either immediately or after an imposed delay lasting from 0.5 to 2 s. In these subjects, digital time stamps corresponding to well entry and reward delivery were recorded and used for illustration in Fig. 3c, but were otherwise not used in determining entry times or occupancy of the subjects at the wells for consistency among all subjects. Task epochs lasting 15 min were preceded and followed by rest epochs lasting ~20 min in a high-walled black box (floor edges 25–35 cm and height 50 cm), during which rats often groomed, quietly waited, and slept. Two subjects also ran in an open field environment for scattered food (grated cheese) after W-maze recordings, with additional interleaved rest epochs. Tetrode positions were adjusted after each day's recordings.

Data were collected using the NSpike data acquisition system (L.M.F. and J. MacArthur, Harvard Instrumentation Design Laboratory). During recording, an infrared diode array with a large and a small cluster of diodes was affixed to headstage preamplifiers to enable tracking of head position and head direction. Following recording, position and direction were reconstructed using a semi-automated analysis of digital video (30 Hz) of the experiment. Spike data were recorded relative to the REF tetrode, sampled at 30 kHz, digitally filtered between 600 Hz and 6 kHz (2-pole Bessel for high- and low-pass), and threshold crossing events were saved to disk. Local field potentials (LFPs) were sampled at 1.5 kHz and digitally filtered between 0.5 Hz and 400 Hz. LFPs analysed were relative to the REF tetrode except where otherwise indicated.

Individual units (putative single neurons) were identified by clustering spikes using peak amplitude, principal components, and spike width as variables (MatClust, M.P.K.). Only well-isolated neurons with stable spike waveform amplitudes were clustered. A single set of cluster bounds defined in amplitude and width space could often isolate units across an entire recording session. In cases where there was a shift in amplitudes across time, units were clustered only when that shift was coherent across multiple clusters and when plots of amplitude versus time showed a smooth shift. No units were clustered in which part of the cluster was cut off at spike threshold.

Histology and recording site assignment. After recordings, subjects were anaesthetized with isoflurane, electrolytically lesioned at each tetrode (30 μA of positive current for 3 s applied to two channels of each tetrode), and allowed to recover overnight. In one subject, no electrolytic lesions were made, and tetrode tracks rather than lesions were used to identify recording sites. Subjects were euthanized with pentobarbital and were perfused intracardially with PBS followed by 4% paraformaldehyde in PBS. The brain was post-fixed *in situ* overnight, after which the tetrodes were retracted and the brain removed, cryo-protected (30% sucrose in PBS), and embedded in OCT compound. Coronal (7 subjects) and sagittal (1 subject) sections (50 μm) were taken with a cryostat. Sections were either Nissl-stained with cresyl violet or stained with the fluorescent Nissl reagent NeuroTrace Blue (1:200) (Life Technologies, N-21479). In four subjects, the sections were blocked (5% donkey serum in 0.3% Triton-X in TBS, used for all incubations) for 1 h, incubated with RGS14 (refs 36, 47, 71) antibody (1:400) (Antibodies Inc., 75-140) overnight, washed, and subsequently incubated with fluorescent secondary antibody (1:400) (Alexa 568, Life Technologies). CA2 recording sites were designated as those in which the electrolytic lesion or end of tetrode track overlapped with the dispersed cytoarchitectural zone characteristic of CA2 (refs 28–30, 47, 50, 54–57). This strategy was deliberately inclusive to maximize detection of putative CA2 neurons with novel physiological responses (N units, Fig. 1 and Extended Data Fig. 3). It is important to note that CA2 sites defined in this way include recording locations that have been designated in previous studies as 'CA3a'.

Data analysis. All analyses were carried out using custom software written in Matlab (Mathworks).

SWR detection. Detection of SWRs was prerequisite for all data analysed in this study, and was performed only when at least three CA1 cell layer recordings were available. Offline, a multisite average approach was used to detect SWRs⁵⁸. Specifically, LFPs from all available CA1 cell layer tetrodes were filtered between 150–250 Hz, then squared and summed across tetrodes. This sum was smoothed with a Gaussian kernel ($\sigma = 4$ ms) and the square root of the smoothed sum was analysed. SWRs were detected when the signal exceeded 2 s.d. of the recording epoch mean for at least 15 ms. SWR periods were then defined as the periods,

containing the times of threshold crossing, in which the power trace exceeded the mean. SWR onset was defined as the start of a SWR period. Detection of SWRs was performed only when subjects' head speed was $<4\text{ cm s}^{-1}$. For SWR-triggered spike raster plots and PSTH plots, a 0.5 s exclusion period was imposed to isolate SWRs occurring only after non-SWR periods; otherwise, analyses of SWRs included all detected SWRs.

Unit inclusion. Two unit sets were analysed in this study. In the first (task unit set), units included fired at least 100 spikes outside of SWRs in at least one task epoch. In the second set (rest unit set), units included fired at least 100 spikes outside of SWRs in at least one rest epoch, moreover specifically in awake periods (see below). The rest unit set was established to evaluate spatial representations and network patterns in the rest environment. For both unit sets, all included units were required to have available data for least 300 (typically $>1,000$) concurrently detected SWRs in either task or rest epochs. Since relatively less is understood about hippocampal neurons in CA2, units recorded at CA2 in the rest unit set were included in the study only if they met the task unit set criterion to ensure that neurons engaged during active behaviour were evaluated. All unit population findings in this study refer to the task unit set, with the exception of those presented in Fig. 5f and Extended Data Fig. 10, which refer to the rest unit set.

Principal versus interneuronal unit classification. For each unit set, scatter plots of firing rate, spike width, and autocorrelation function mean (calculated from 0 to 40 ms; low values indicating burst firing) showed two distinct clusters^{18,31,59–62} (example plot of task unit set in Fig. 1c). Putative principal units corresponded with the low firing rate ($<4\text{ Hz}$), large spike width, low autocorrelation mean cluster, while putative interneuronal units corresponded to the cluster characterized by high firing rate, small spike width, and high autocorrelation mean. Twenty-one units with ambiguous features were left unclassified. All units in the study were isolated (clustered) and classified before STA analysis.

N versus P unit classification. Periods when head speed was $<4\text{ cm s}^{-1}$ were segregated into SWR versus non-SWR periods, and the change in firing rate during SWRs calculated. The period types were then permuted ($n=1,000$) to obtain a distribution of firing rate differences given the null hypothesis of no association of firing rate with period type. P units were those units showing a difference in firing rate that was $>95\%$ of values from the null distribution, either for SWRs of any single task epoch or for rest epoch SWRs. N units were those that showed a failure of significance for SWRs in every task epoch and also for rest epoch SWRs. This approach minimized false positives in the detection of N units. Negatively modulated (inhibited) units were formally identified as a subset of N units (examples in Fig. 1c, d and additional observations in Extended Data Fig. 3b) showing a firing rate difference during SWRs that was $<95\%$ of the values from the null hypothesis distribution for rest epoch SWRs and also for SWRs of at least one task epoch.

A small number of CA1 principal units (11 out of 504) and CA3 principal units (7 out of 289) were classified as N units (N versus P proportions for the task unit set shown in Fig. 1e); these units were excluded from all analyses. After exclusion of N units for CA1 and CA3, total putative principal unit counts in the task unit set were CA1: 478, CA3: 271, CA2 P: 142, CA2 N: 84; in the rest unit set, CA1: 163, CA3: 76, CA2 P: 76, CA2 N: 68. Throughout this study, 'N units' and 'P units' solely refer to the distinct unit populations recorded at CA2 sites, and are equivalent to 'CA2 N' and 'CA2 P'.

Behavioural state. Periods of locomotion were defined as times when head speed was $>4\text{ cm s}^{-1}$. Periods of non-SWR immobility were times when head speed was $<4\text{ cm s}^{-1}$ separated from locomotor periods by 2 s buffer intervals (preceding and following) and excluding SWR periods. Thus brief interruptions in locomotion did not qualify as formally detected periods of immobility.

Firing rate estimation. For each unit, instantaneous firing rate (IFR) was estimated by convolving the unit's spike train (1-ms bins) with a Gaussian kernel ($\sigma=250\text{ ms}$). Mean firing rates in the task (Fig. 2c and Extended Data Fig. 8b) were calculated from the task epoch in which the unit had the highest mean firing rate combined with additional task epochs of the same environment (specific W-maze) when available. Mean firing rates in the rest environment (Fig. 5c) were calculated from all available rest epochs, and were only calculated for units for which LIA and SIA sleep data were available. Firing rates during SWRs were calculated for SWR periods in either task epochs (Fig. 2c and Extended Data Fig. 8b) or rest epochs (Fig. 5c).

Firing versus speed correlation. For each unit, the Pearson correlation coefficient (r) was calculated between IFR and the logarithm of head speed^{27,63,64} for non-SWR periods. The correlation was calculated from the task epoch in which the unit had the highest mean firing rate combined with additional task epochs of the same W-maze when available. Only units with significant correlations ($P<0.05$) were analysed (CA1: 475/477 units, CA2 P: 141/142 units, CA2 N: 83/84 units, CA3: 270/271 units). It is worth noting that the findings relating CA2 N unit firing to speed in the task condition (Fig. 2) are not a direct consequence of the N

unit classification criteria, which refer strictly to a lack of increased firing during SWRs.

Spatial firing. To quantify spatial coverage, 2D position data (corresponding to subjects' head location) for all subjects was first converted to linear position. Linear position was measured as the distance from the centre reward well along the linear arms of the W-shaped task maze. In addition, all linear positions were classified as belonging to one of four possible trajectories of the behavioural task, namely, outbound and inbound trajectories between the centre well and each of the two outer wells (diagrammed in Extended Data Fig. 1a). The end of each continuous trajectory assignment period corresponded to the separation of the subject's linear position from that of the target well of the given trajectory ($>2\text{ cm}$ from well).

No trajectory assignment was performed for periods of data corresponding to three cases: (1) excursions in which the subject departed and returned to the same well, (2) excursions in which the subject occupied a maze segment that was not part of the three linear segments defining the animal's current trajectory, and (3) times during which the subject's linearized head direction (either forward or backward along the current maze segment) did not match the defined direction of the animal's current trajectory. These unassigned periods represented a minority proportion of the data (33% across all task sessions) and were not included either in spatial plots referencing trajectory (occupancy-normalized firing maps in Extended Data Fig. 4b) or in subsequent spatial coverage analysis, which relied on unambiguous trajectory assignment in accordance with known direction- and trajectory-dependence of hippocampal spatial firing^{25,65–67}. Less stringent restriction of positional data produced qualitatively equivalent results.

For each unit, an occupancy-normalized firing map was calculated for each of the four task trajectories. First, total spike counts and occupancy durations were calculated for 2-cm spatial bins on each trajectory. Both the occupancy and spike counts per bin were smoothed with a Gaussian ($\sigma=4\text{ cm}$), then spike counts were divided by occupancy to produce the unit's smoothed occupancy-normalized firing map. The peak spatial firing rate was the maximum value in the occupancy-normalized map. A bin counted towards spatial coverage (Fig. 3b) if its occupancy-normalized rate was $>2\text{ Hz}$. Spatial coverage was quantified in each unit's highest mean firing rate task epoch. Seven units (CA1: 2 units, CA2 N: 5 units) were not included in spatial coverage quantification because of a failure of subjects to visit one of the maze arms in the units' highest firing rate task epochs. Quantification using additional velocity cutoffs and spatial firing thresholds is shown in Extended Data Fig. 4a.

Two-dimensional occupancy-normalized firing maps were constructed with 1-cm (W-maze) or 0.5-cm (rest environment) square bins. For example plots, these maps were smoothed with a symmetric 2D Gaussian ($\sigma=3\text{ cm}$ for maze; $\sigma=1.5\text{ cm}$ for rest environment); for nesting position analyses in the rest environment, no smoothing was performed. Data during SWR periods were excluded from all spatial firing plots and analyses.

Well firing. Well periods were defined as times when the subject's linear position matched that of the reward well ($<2\text{ cm}$ separation). Well visits were defined as well periods that lasted at least 2 s and were preceded earlier in the recording epoch by a well period at a different well. In instances in which subjects re-visited the well they departed from before visiting another well, a well visit was only registered after an exclusion period of 5 s. Well entry times (designated $t=0$ in well raster plots) were defined as the beginning of well visits.

To calculate the well specificity index (WSI) of a unit, the well firing rate at each of the three wells of the task was first determined. Well firing rate was specifically calculated from the intersection of well periods with non-SWR immobility periods (well intersectional time). Next, each of the three well firing rates was divided by the numerical sum of the three well firing rates (normalization) to create a three-category (well A versus B versus C) probability distribution of firing activity. This probability distribution was subsequently treated as a circular distribution with a vector whose length corresponded to the probability mass for well A placed at 0° , a vector for well B at 120° , and a vector for well C at 240° . The magnitude of the vector sum (resultant), defined as the WSI, was used as a measure of well-specific firing. The WSI directly reflects specificity of firing: a WSI = 0 corresponds to equal firing at all three wells (completely non-specific), WSI = 0.5 corresponds to firing at two wells, and WSI = 1 corresponds to firing at one well.

The WSI was calculated in a unit's highest mean firing task epoch, and was only calculated when (i) at least 100 spikes were observed during well intersectional time, (ii) at least 5 s of well intersectional time was available for each of the three wells, (iii) the firing rate (during well intersectional time) for at least one well exceeded 0.5 Hz. These minimum activity criteria ensured that the WSI was calculated only for units that were unequivocally active at wells and for which adequate data at each well were available.

Theta analysis. To estimate theta phase, LFP from the REF tetrode (located in corpus callosum overlying right dorsal hippocampus⁶⁸) was filtered at 5–11 Hz. The phase of the Hilbert transform of the filtered REF LFP was then designated as

the theta phase^{31,68}. For a given unit, theta phase locking analysis was performed for locomotor periods ($> 4 \text{ cm s}^{-1}$) in task epochs, and moreover only when at least 50 spikes were present in these periods.

Spike- and SWR-triggered averaging of LFP (STA and RTA). Spike-triggered averages of LFP (STAs) were calculated for spiking in task epochs, moreover specifically for two distinct period types: locomotion and non-SWR immobility. For a given unit, STAs were calculated only when at least 100 spikes in the period type were observed. In each subject, the recording electrodes for each of four LFP reference regions (REF and CA2, CA3, and DG when available) were kept constant over all recording days. Each LFP recording site either reported principal units for its correspondent region (if CA2, CA3, DG) or was within 60 μm of the depth range at which principal units were detected, as determined from records of tetrode adjustment depths. In cases where the LFP reference region was the same as the region in which the unit was located, the parent electrode of the unit was chosen as the LFP reference.

For each unit for which an STA was calculated, a matched SWR-triggered average of LFP (RTA) was calculated, using the same LFP reference site and averaging across all SWRs detected in the same task recording epochs as the unit. RTAs were calculated by averaging LFP aligned to the time of peak power (designated $t=0$) in the multisite ripple band power (power at 150–250 Hz across CA1 sites, see above) for each SWR.

To evaluate the spectral components of the STAs and RTAs, the power spectral density (PSD) of individual unit STAs and RTAs (2-s LFP traces) was calculated using Welch's method (pwelch, Matlab Signal Processing Toolbox). Spectral analysis is shown for STAs/RTAs of LFP recorded in DG (Fig. 4d), as DG LFP showed the largest amplitude low-frequency signals.

N wave firing. To detect unit firing in association with the N wave, unit STAs were analysed. Specifically, unit STAs were classified into distinct groups using the following procedure. First, non-SWR immobility STAs and RTAs were calculated from LFP filtered at 1–4 Hz. Since the N wave as originally identified (Fig. 4c) was largest at DG, then CA3, and then CA2, the STAs were calculated for LFP at DG sites when available, then at CA3 when available, then at CA2. Furthermore, for an LFP recording site to be used to calculate classifiable STAs, the RTA at that site had to be significantly negative at $t=0$ ($P < 0.001$ level, signed-rank). In a small number of cases in which this condition was not satisfied, LFP from the next available region, if available, was used. Thus SWR sharp waves were verified to manifest as negative deflections at recording sites used to calculate STAs.

A unit STA was classified in two specific cases: (1) when the STA at the time of spiking ($t=0$) was positive and the nearest local extremum was a maximum (peak), and (2) when the STA at the time of spiking was negative and nearest local extremum was a minimum (trough). A small number of units showing positive troughs or negative peaks were left unclassified (CA1: 10 out of 146 units, CA2 N units: 1 out of 58 units, CA3: 3 out of 137 units, interneurons: 10 out of 63 units, plotted at bottom in Extended Data Figs 7b, 7d and 8a). Units satisfying (1) and (2) are referred to as 'positive STA' and 'negative STA' unit populations, respectively. Units satisfying (1) were identified as firing in association with the N wave (N wave-coupled).

Sleep state identification. In rest epochs, awake periods were identified as times in which head speed was $> 4 \text{ cm s}^{-1}$ in addition to times $< 4 \text{ cm s}^{-1}$ within 7 s of a previous movement $> 4 \text{ cm s}^{-1}$. Thus, given the behavioural state criteria (see above), for each distinct period in which a subject stopped moving, no more than 5 s were included as awake immobility.

Candidate sleep periods were identified as times $< 4 \text{ cm s}^{-1}$ preceded by 60 s with no movement $> 4 \text{ cm s}^{-1}$. REM periods within candidate sleep times were identified following an established procedure⁶⁹. Specifically, the ratio of Hilbert amplitudes (smoothed with a Gaussian kernel, $\sigma = 1 \text{ s}$) of theta (5–11 Hz) to delta (1–4 Hz) filtered LFP was calculated for all available CA1 tetrodes (referenced to cerebellar ground), and the mean taken over tetrodes. For each rest epoch, a threshold (range: 1.2–1.8) was manually set to capture sustained periods (10 s minimum duration) in which the theta:delta ratio was elevated. LFP and position data from each detected REM period were visually inspected.

For a given day's set of candidate sleep times outside of REM periods, LFP from each available CA1, CA3, and DG recording site was squared then smoothed with a Gaussian kernel ($\sigma = 300 \text{ ms}$). The square root of the smoothed signal was then z-scored and summed across sites. The sum trace was in turn z-scored to obtain an aggregate hippocampal LFP amplitude. For each rest epoch, the distribution of aggregate LFP amplitudes was plotted (example trace and distribution in Fig. 5a). From a rest epoch in which bimodality was observed, the value at the local minimum separating the two modes was chosen as the SIA z-score threshold for the day. SIA periods were defined as non-REM times in which the aggregate LFP amplitude was below the threshold, and LIA periods otherwise. In a minority of cases, a threshold was chosen to isolate a heavy left tail of the distribution, later verified in the LFP to correspond to SIA periods. Across all recording days ($n = 73$

days) the SIA threshold was -0.67 ± 0.24 (z-score, mean \pm s.d.), and median period durations were SIA: 1.20 s; LIA: 2.48 s; REM: 27 s. Visual inspection of LFPs confirmed that SIA periods could often be $\sim 1 \text{ s}$ in duration³³, indicating rapid switching between distinct sleep states (Fig. 5a, b). Also, as previously reported³³, slight movements without overt awaking could at times be observed during SIA (Fig. 5b). Lastly, though SWRs in sleep typically occurred during LIA, SWRs at times occurred within identified SIA periods³³. Thus, to isolate SIA periods optimally, SWR periods were not included in calculations referencing SIA periods.

Sleep periods were candidate sleep periods at least 90 s in duration and containing extended ($> 5 \text{ s}$) continuous LIA periods. Across all recording days, 465 sleep periods (median duration: 218 s) were identified.

Nesting position coding. Unit firing rates during SIA were calculated for individual sleep periods. Sleep periods in which a unit's SIA firing rate was $> 2 \text{ Hz}$ were categorized as SIA ON for the unit, and SIA OFF otherwise. Next, the 2D spatial firing map (non-smoothed, see above) for the unit from awake periods in the same ~ 20 -min rest epoch was referenced. During awake periods, the total number of spikes and total time spent at positions $> 5 \text{ cm}$ from the subject's head position at the beginning of the sleep period (nesting position) were categorized as Nest OUT, and likewise Nest IN for positions $< 5 \text{ cm}$. If there were additional sleep periods of a given type (SIA ON or SIA OFF) available for a unit, then the spike counts and durations spent were summed within the Nest OUT/IN categories for the respective nesting positions of the additional sleep periods. Firing rate for a given category (for example, SIA ON, Nest OUT) was calculated as the total number of spikes divided by the total time.

A unit coding for nesting position is expected to show two firing patterns (dual criteria): if classified as SIA ON in a given sleep period, the unit is expected to show higher firing rates, during awake periods, at positions nearer to the nesting position (Nest IN, $< 5 \text{ cm}$) corresponding to the sleep period; conversely, if classified as SIA OFF in a given sleep period, the unit is expected to show higher firing rates, during awake periods, at positions farther from the nesting position (Nest OUT, $> 5 \text{ cm}$) corresponding to the sleep period.

Unit populations were tested for nesting position coding with two approaches. In the first, absolute firing rates were compared between Nest IN versus OUT periods for both SIA ON and SIA OFF groups³³ (Extended Data Fig. 10j). In the second (Fig. 5f), firing rates in the Nest IN versus Nest OUT conditions were compared for each unit by calculating a measure termed the nesting position specificity index, calculated as $2 \times f_{\text{IN}} / (f_{\text{IN}} + f_{\text{OUT}}) - 1$. Using this measure, a firing rate in Nest IN that is twice as high as in Nest OUT yields a value of 1/3; three times as high yields a value of 1/2.

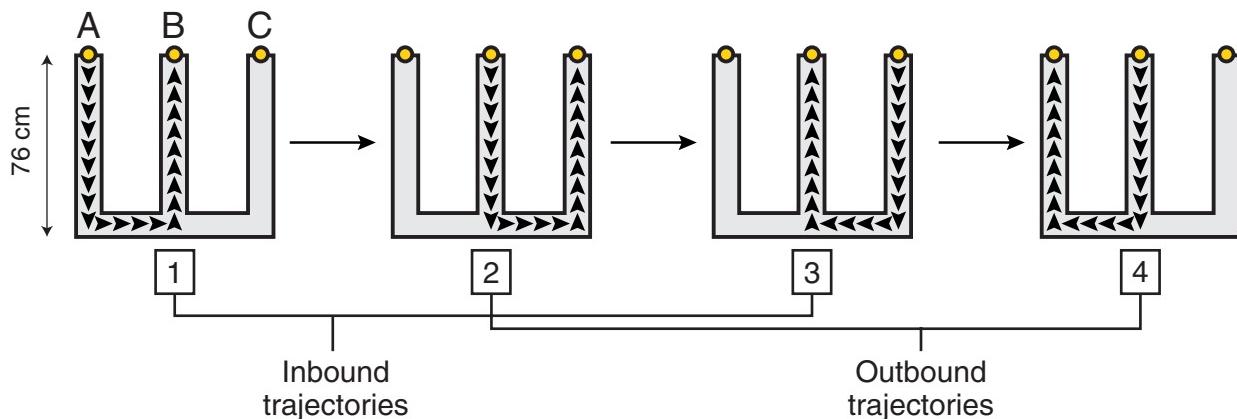
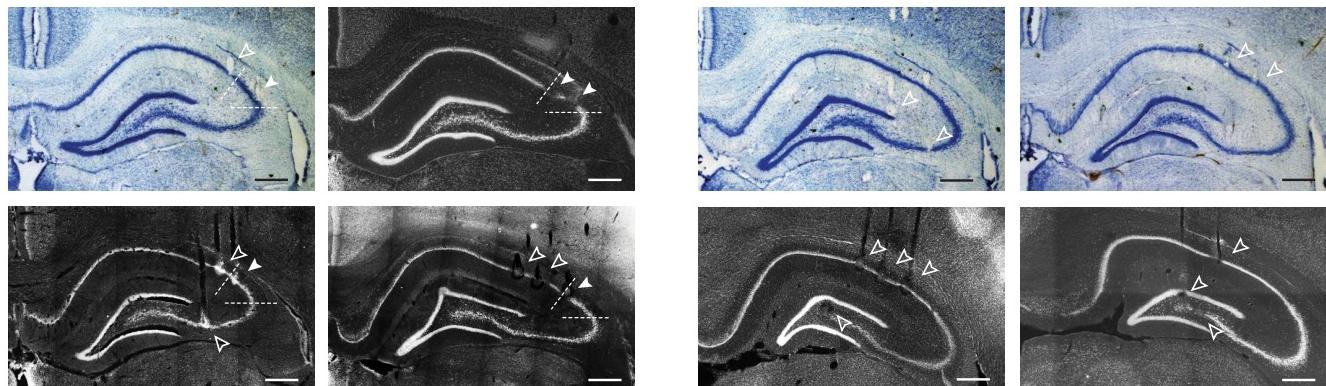
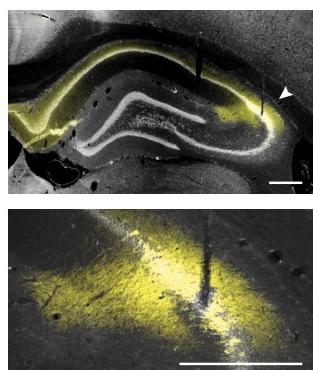
For either the absolute firing rate or the specificity index approach, the dual criteria for nesting position coding in a unit population were (1) higher firing during Nest IN versus Nest OUT for the SIA ON group and (2) higher firing during Nest OUT versus Nest IN for the SIA OFF group.

Statistics. All statistical tests were two-sided.

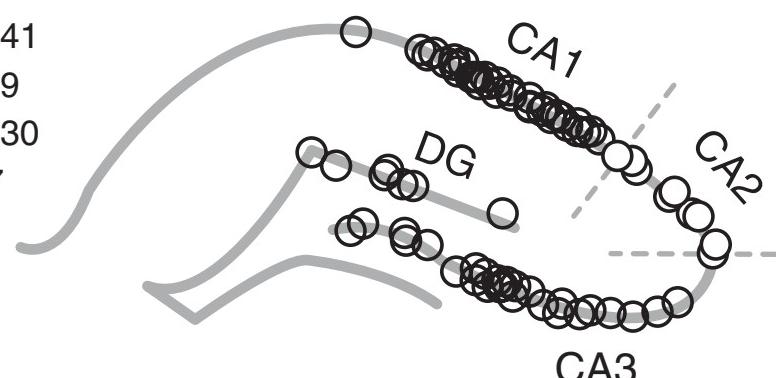
Code availability. All custom-written code is available upon request.

51. Gray, C. M., Maldonado, P. E., Wilson, M. & McNaughton, B. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J. Neurosci. Methods* **63**, 43–54 (1995).
52. Potter, G. B., et al. Generation of Cre-transgenic mice using *Dlx1/Dlx2* enhancers and their characterization in GABAergic interneurons. *Mol. Cell Neurosci.* **40**, 167–186 (2009).
53. Neunuebel, J. P. & Knierim, J. J. Spatial firing correlates of physiologically distinct cell types of the rat dentate gyrus. *J. Neurosci.* **32**, 3848–3858 (2012).
54. Ishizuka, N., Cowan, W. M. & Amaral, D. G. A quantitative analysis of the dendritic organization of pyramidal cells in the rat hippocampus. *J. Comp. Neurol.* **362**, 17–45 (1995).
55. Woodhams, P. L., Celio, M. R., Ulfhake, N. & Witter, M. P. Morphological and functional correlates of borders in the entorhinal cortex and hippocampus. *Hippocampus* **3**, 303–311 (1993).
56. Amaral, D. G. & Lavenex, P. in *The Hippocampus Book* (eds Andersen, P. et al.) 37–114 (Oxford Univ. Press, 2007).
57. Cui, Z., Gerfen, C. R. & Young, W. S., III. Hypothalamic and other connections with dorsal CA2 area of the mouse hippocampus. *J. Comp. Neurol.* **521**, 1844–1866 (2013).
58. Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A. & Buzsaki, G. Fast network oscillations in the hippocampal CA1 region of the behaving rat. *J. Neurosci.* **19**, RC20 (1999).
59. Ranck, J. B. Jr. Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. I. Behavioral correlates and firing repertoires. *Exp. Neurol.* **41**, 462–531 (1973).
60. Fox, S. E. & Ranck, J. B. Jr. Electrophysiological characteristics of hippocampal complex-spike cells and theta cells. *Exp. Brain Res.* **41**, 399–410 (1981).
61. Skaggs, W. E. & McNaughton, B. L. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* **271**, 1870–1873 (1996).

62. Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A. & Buzsaki, G. Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *J. Neurosci.* **19**, 274–287 (1999).
63. Kemere, C., Carr, M. F., Karlsson, M. P. & Frank, L. M. Rapid and continuous modulation of hippocampal network state during exploration of new places. *PLoS ONE* **8**, e73114 (2013).
64. Chen, Z., Resnik, E., McFarland, J. M., Sakmann, B. & Mehta, M. R. Speed controls the amplitude and timing of the hippocampal gamma rhythm. *PLoS ONE* **6**, e21408 (2011).
65. Frank, L. M., Brown, E. N. & Wilson, M. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* **27**, 169–178 (2000).
66. Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
67. Ito, H. T., Zhang, S. J., Witter, M. P., Moser, E. I. & Moser, M. B. A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. *Nature* **522**, 50–55 (2015).
68. Lubenov, E. V. & Siapas, A. G. Hippocampal theta oscillations are travelling waves. *Nature* **459**, 534–539 (2009).
69. Mizuseki, K., Diba, K., Pastalkova, E. & Buzsaki, G. Hippocampal CA1 pyramidal cells form functionally distinct sublayers. *Nature Neurosci.* **14**, 1174–1181 (2011).
70. JadHAV, S. P., Kemere, C., German, P. W. & Frank, L. M. Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336**, 1454–1458 (2012).
71. Lee, S. E. et al. RGS14 is a natural suppressor of both synaptic plasticity in CA2 neurons and hippocampal-based learning and memory. *Proc. Natl Acad. Sci. USA* **107**, 16994–16998 (2010).
72. Suzuki, S. S. & Smith, G. K. Spontaneous EEG spikes in the normal hippocampus. I. Behavioral correlates, laminar profiles and bilateral synchrony. *Electroencephalogr. Clin. Neurophysiol.* **67**, 348–359 (1987).
73. Buzsáki, G. Two-stage model of memory trace formation: a role for “noisy” brain states. *Neuroscience* **31**, 551–570 (1989).
74. Harris, K. D., Hirase, H., Leinekugel, X., Henze, D. A. & Buzsaki, G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**, 141–149 (2001).
75. Mizuseki, K., Royer, S., Diba, K. & Buzsaki, G. Activity dynamics and behavioral correlates of CA3 and CA1 hippocampal pyramidal neurons. *Hippocampus* **22**, 1659–1680 (2012).
76. Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).
77. Epsztain, J., Brecht, M. & Lee, A. K. Intracellular determinants of hippocampal CA1 place and silent cell activity in a novel environment. *Neuron* **70**, 109–120 (2011).
78. Grienberger, C., Chen, X. & Konnerth, A. NMDA receptor-dependent multidendrite $\text{Ca}(2+)$ spikes required for hippocampal burst firing *in vivo*. *Neuron* **81**, 1274–1281 (2014).
79. Bitner, K. C. et al. Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. *Nature Neurosci.* **18**, 1133–1142 (2015).
80. Klausberger, T. & Somogyi, P. Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* **321**, 53–57 (2008).
81. Diba, K., Amarasingham, A., Mizuseki, K. & Buzsaki, G. Millisecond timescale synchrony among hippocampal neurons. *J. Neurosci.* **34**, 14984–14994 (2014).
82. Royer, S. et al. Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. *Nature Neurosci.* **15**, 769–775 (2012).
83. Skaggs, W. E., McNaughton, B. L., Gothard, K. & Markus, E. in *Advanced in Neural Information Processing Systems* (eds Hanson, S., Cowan, J. D. & Giles, C. L.) 1030–1037 (Morgan Kaufmann, 1993).

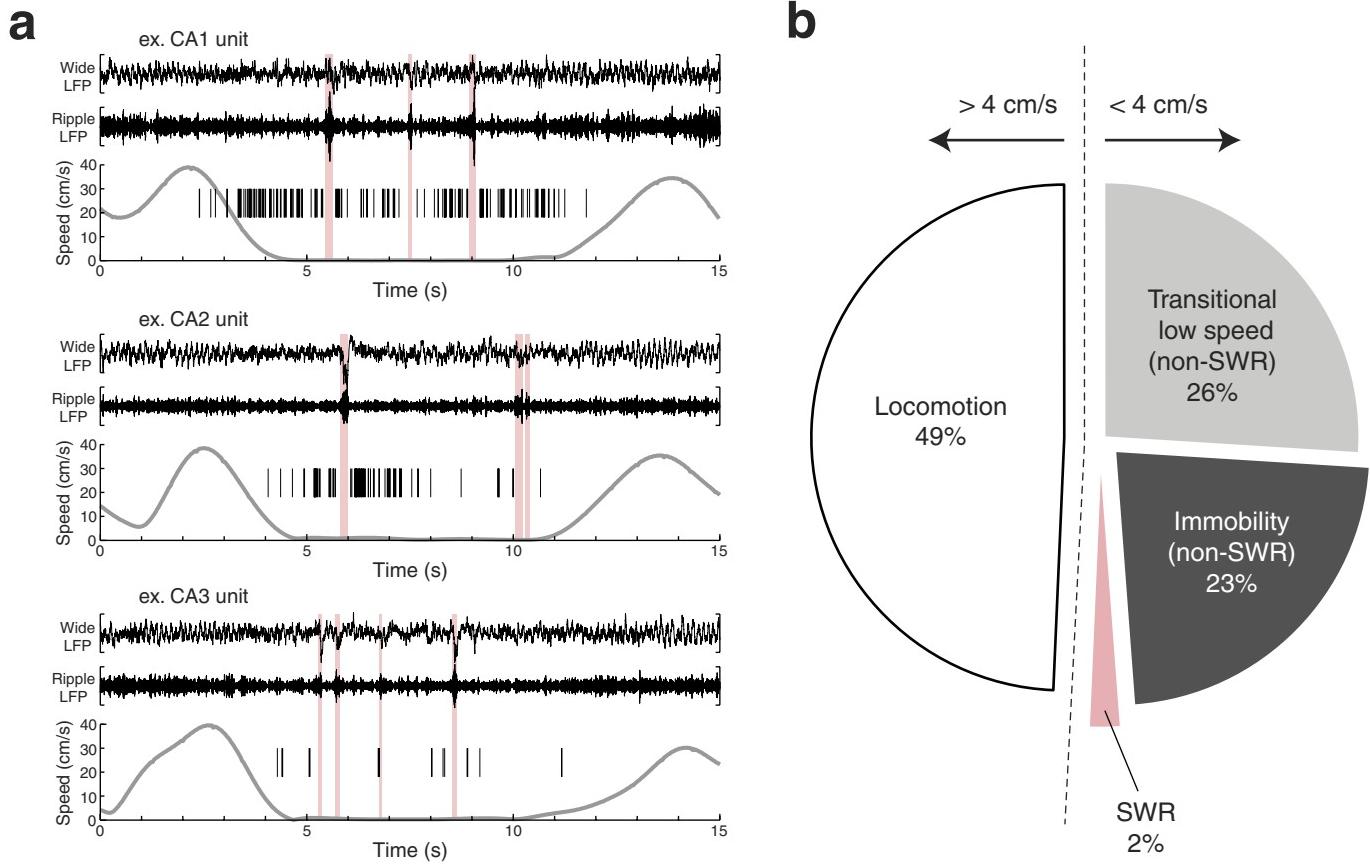
a**b****d****e**

CA1: 41
CA2: 9
CA3: 30
DG: 7



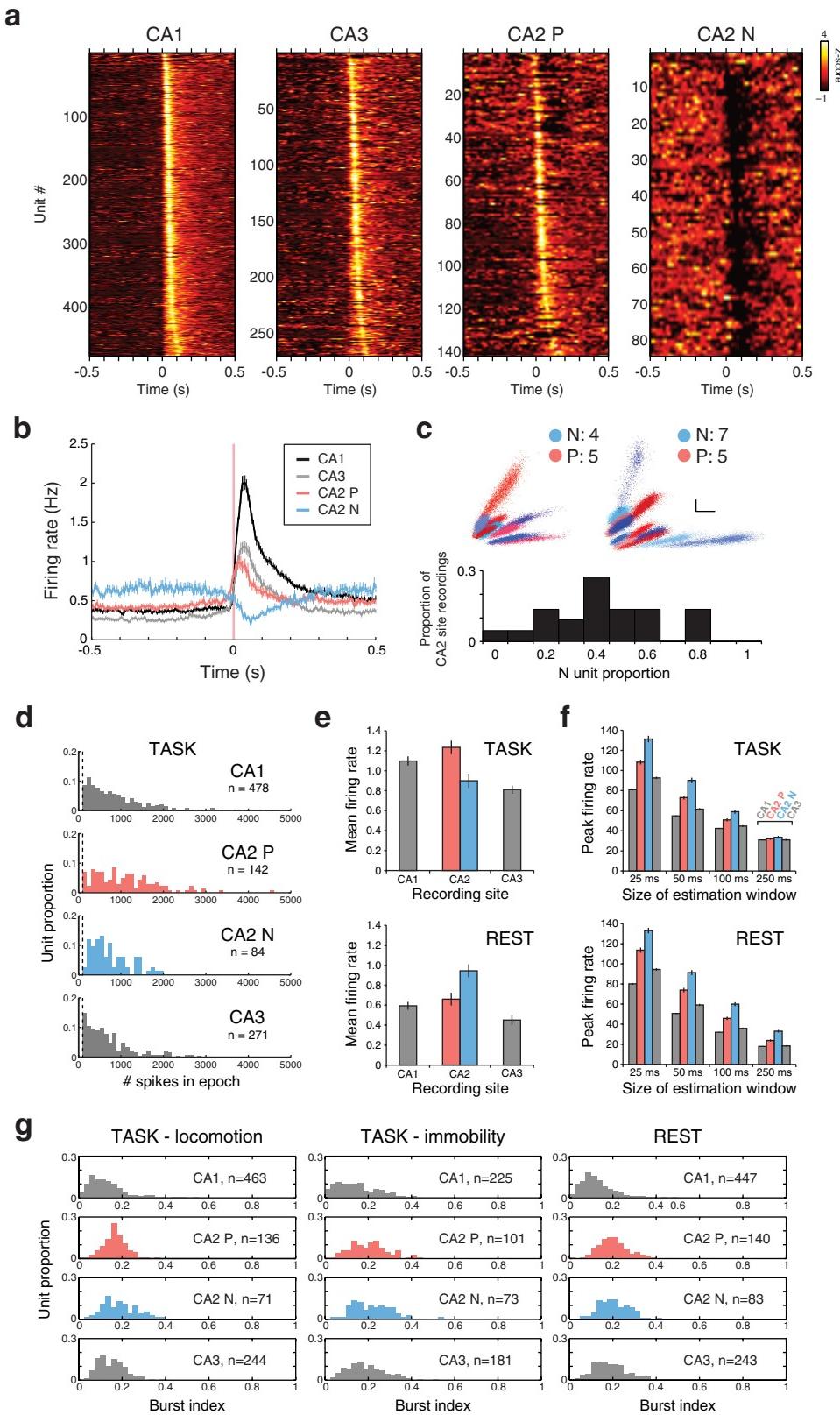
Extended Data Figure 1 | Behavioural task and hippocampal recording sites. **a**, Continuous spatial alternation task^{21,24,65,70}. The task environment is a W-shaped maze with a centre arm and two outer arms. Reward (~0.3 ml of sweetened evaporated milk) is dispensed through 3-cm diameter wells (yellow circles; designated 'A', 'B', and 'C' for reference in data plots), located at the end of each arm. Rats are rewarded for performing the trajectory sequence shown (numbered 1–4), in which the correct destination after visiting the centre well is the less recently visited outer well. All subjects stopped locomoting upon reaching the reward wells to check for (by licking) and consume reward. Subjects also stopped intermittently elsewhere in the maze (most frequently at maze junctions), particularly in earlier exposures to the task. **b, c**, Example hippocampal histological sections showing tetrode tracks and electrolytic lesions in CA1, CA2, CA3, and DG. Nissl-stained sections show neuronal cell bodies in dark blue, while sections stained with Neurotrace show neuronal cell bodies in light grey. Panel **b** shows example sections with sites overlapping with the CA2 cytoarchitectural locus^{28–30,36,47,50,54–57}.

(enclosed by dotted lines; characterized by dispersion of the hippocampal cell layer in the region between CA1 and CA3). Filled arrowheads indicate sites overlapping with CA2, while empty arrowheads indicate non-CA2 recording sites. The CA2 site assignment was deliberately inclusive to maximize detection of units at CA2 with novel physiological responses (N units, Fig. 1 and Extended Data Fig. 3). Scale bars: 500 µm. **d**, Coronal hippocampal section stained with a neuronal cell body marker (light grey; NeuroTrace) and CA2 marker (yellow; RGS14^{36,47,71}). Bottom, magnified view of a track left by a CA2 site tetrode. Scale bars: 500 µm. **e**, Survey of recording sites included in the study data set. Diagrammed in a representative hippocampal section are recording site locations (circles) of seven subjects from which coronal hippocampal sections were taken (CA1: 41 sites, CA2: 9 sites, CA3: 30 sites, DG: 7 sites; two additional CA2 sites near the septal pole of hippocampus not shown). Dotted lines enclose the CA2 anatomical locus, with overlapping recording sites shown as filled circles. The majority of CA1 recordings were in CA1c, while the majority of CA3 recordings were in CA3b.



Extended Data Figure 2 | Observation of firing during immobility.
a, Non-SWR immobility firing in three example principal units recorded in CA1, CA2, and CA3. Each firing raster is shown as vertical lines overlaid on a plot of the subject's head speed (grey trace). Top traces: wide-band LFP (0.5–400 Hz, scale bar: 800 μ V) and ripple-band LFP (150–250 Hz, scale bar: 100 μ V) traces from a simultaneous recording in CA1, to show hippocampal network state. SWR periods are plotted as pink zones. Note that substantial firing occurs in the absence of (i) locomotion, (ii) detectable SWRs, and (iii) detectable theta (regular ~8 Hz rhythm visible in the LFP during moving periods). **b**, Proportions of time spent

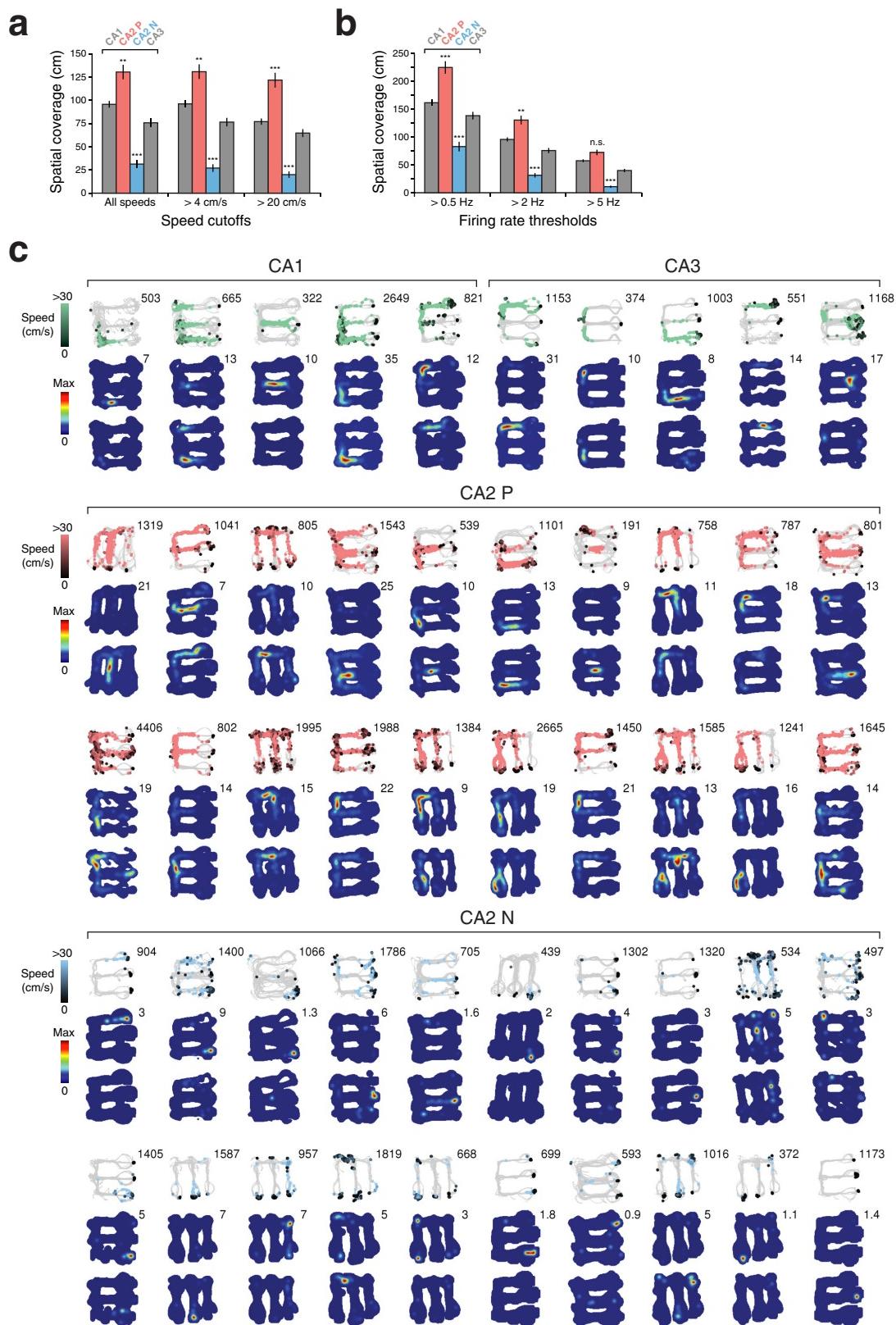
in different period types over all task recording epochs ($n = 222$ task recording epochs, 8 subjects) in the data set. During the performance of the task, a substantial proportion of time was spent at low speeds and immobility, moreover when SWRs were not detected. Transitional low speed periods were times when the subject's speed was $< 4 \text{ cm s}^{-1}$ and within 2 s (earlier or later) of periods of movement $> 4 \text{ cm s}^{-1}$, while immobility periods were times when the speed was $< 4 \text{ cm s}^{-1}$ and separated more than 2 s (earlier or later) from periods of movement $> 4 \text{ cm s}^{-1}$. Note that SWR periods comprised only a minority of time spent at low speeds, consistent with past observations^{17,72,73}.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Firing properties of CA1, CA2, and CA3 units. **a**, Peri-SWR time histograms (PSTHs; SWR onset at $t = 0$) of firing for all principal units in the task unit set. SWRs from both task and rest epochs were used to calculate PSTHs (1-ms bins), which were smoothed with a Gaussian kernel ($\sigma = 10$ ms). Each unit's mean PSTH was then z -scored (colour bar) and plotted in a row. Units are sorted by the time of the maximum z -scored rate from 0 to +100 ms. **b**, PSTHs for the four hippocampal unit populations (mean \pm s.e.m.; number of units: CA1: 478 units; CA3: 271; CA2 P: 142; CA2 N: 84) analysed in this study. Using formal criteria (described in Methods), units that were inhibited during SWRs constituted a majority subset (56 of 84) of N units, and were observed in every subject with CA2 site recordings (5 subjects, inhibition apparent in examples in Fig. 1d and N unit PSTHs in **a**). Here, the reduction of firing in these neurons manifests in the N unit population response as a dip in firing rate at the time of SWRs (N unit population in blue), in contrast to the CA1, CA3, and CA2 P unit populations, all of which showed sharp increases in firing during SWRs¹⁹. Time bins: 5 ms. **c**, Proportion of N units in CA2 site recordings. Upper plots: spike amplitudes measured on two channels of a tetrode for two example CA2 site recordings (left and right). Colours indicate spikes of N (blue-based tones) and P (red-based tones) units. The number of well-isolated principal units of each type is reported at upper right. Scale bars (x and y), 100 μ V. Lower plot: proportion of N units across CA2 site recordings with at least four clustered putative principal units. CA2 recording sites typically reported N and P units concurrently, indicating that the spiking of two distinct hippocampal principal cell types was detectable at a single CA2 recording site. **d**, Unit spike counts in 15-min task epochs for each principal unit population. The counts were taken from each unit's highest mean rate task epoch. Spikes that occurred during SWR periods were

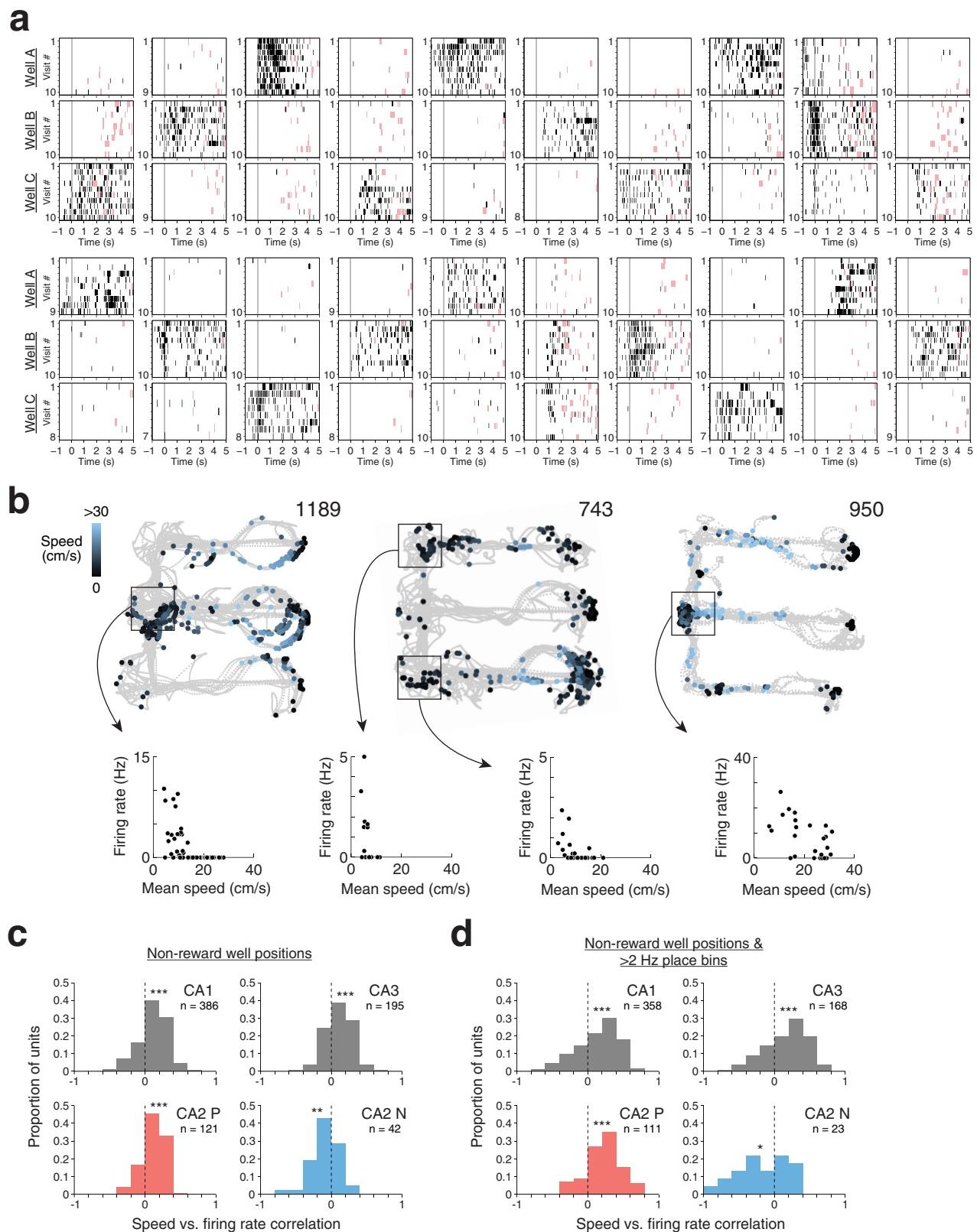
not included in these counts. **e**, Mean firing rate for each principal unit population (mean \pm s.e.m.). The mean rates were calculated from the highest rate epoch for each unit, either among task (top, TASK) or rest (bottom, REST) epochs. TASK number units (task unit set): CA1: 478 units; CA2 P: 142; CA2 N: 84; CA3: 271. REST number units (subset of task unit set with available rest epoch data): CA1: 454 units; CA2 P: 142; CA2 N: 84; CA3: 252. All spikes and epoch times were included. **f**, Peak firing rate for each principal unit population (mean \pm s.e.m.). The peak rates were estimated from the highest rate epochs for each unit, either among task (top, TASK) or rest (bottom, REST) epochs. The peak rate was the maximum instantaneous firing rate (IFR) exhibited by the unit. Here, the IFR was estimated by convolving each unit's spike train (1-ms bins) with Gaussian kernels of different sizes (x -axis, times refer to s.d. of the kernel). TASK number units (task unit set): CA1: 478 units; CA2 P: 142; CA2 N: 84; CA3: 271. REST number units (subset of task unit set with available rest epoch data and at least 100 spikes in a rest epoch): CA1: 421, CA2 P: 138, CA2 N: 82, CA3: 197 units. All spikes and epoch times were included. **g**, Burst firing in each principal unit population. The burst index of a unit was defined as the proportion of inter-spike intervals (ISI) less than 6 ms^{74,75}. Burst indices were calculated separately for three conditions: locomotion (left panels) and immobility (centre) in task epochs, and also for rest epochs (right). In a given condition, a minimum of 100 spikes was required for a unit to be analysed. Moreover, for locomotor and immobility periods from task epochs, only ISIs of spikes that were successive within single uninterrupted periods of a given type were included. Lastly, in this analysis, SWR periods were not excluded. Notably, CA2 N units showed high levels of bursting, suggesting that these units correspond to hippocampal principal (pyramidal) neurons^{59,60,62,76–79}.



Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Spatial firing of CA1, CA2, and CA3 units. For the analyses in **a** and **b**, unit sample sizes are the same as in Fig. 3b. **a**, Spatial coverage at different speed cutoffs (mean \pm s.e.m.), in which only data from periods satisfying the speed condition were analysed. For each speed cutoff, a firing rate threshold of 2 Hz was used. The all speeds condition is the same as in Fig. 3b. CA2 $P >$ each other unit population, Kruskal–Wallis ANOVA, Tukey's post hoc tests, $P = 0.0015$ for all speeds, $P = 0.0021$ for $>4\text{ cm s}^{-1}$, and $P < 10^{-5}$ for $>20\text{ cm s}^{-1}$. CA2 $N <$ each other unit population, Kruskal–Wallis ANOVA, Tukey's post hoc tests, $P < 10^{-6}$ for all speeds, $P < 10^{-7}$ for $>4\text{ cm s}^{-1}$, and $P < 10^{-8}$ for $>20\text{ cm s}^{-1}$. ** $P < 0.01$; *** $P < 0.001$. **b**, Spatial coverage at different firing rate thresholds (mean \pm s.e.m.). For each threshold level, spikes at all speeds were analysed. CA2 $P >$ each other unit population, Kruskal–Wallis ANOVA, Tukey's post hoc tests, $P < 10^{-5}$ for $>0.5\text{ Hz}$, $P = 0.0015$ for $>2\text{ Hz}$, and $P = 0.11$ for $>5\text{ Hz}$. CA2 $N <$ each other unit population, Kruskal–Wallis ANOVA, Tukey's post hoc tests, $P < 10^{-4}$ for $>0.5\text{ Hz}$, $P < 10^{-6}$ for $>2\text{ Hz}$, and $P < 10^{-7}$ for $>5\text{ Hz}$. ** $P < 0.01$; *** $P < 0.001$,

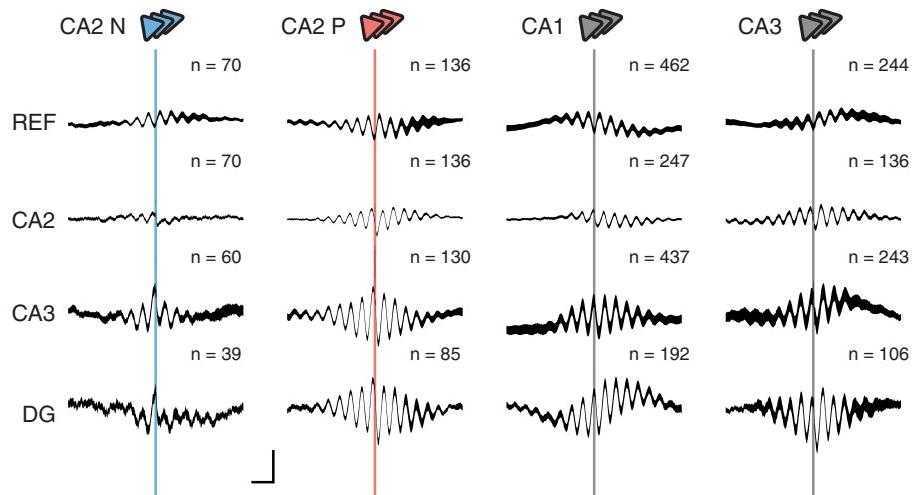
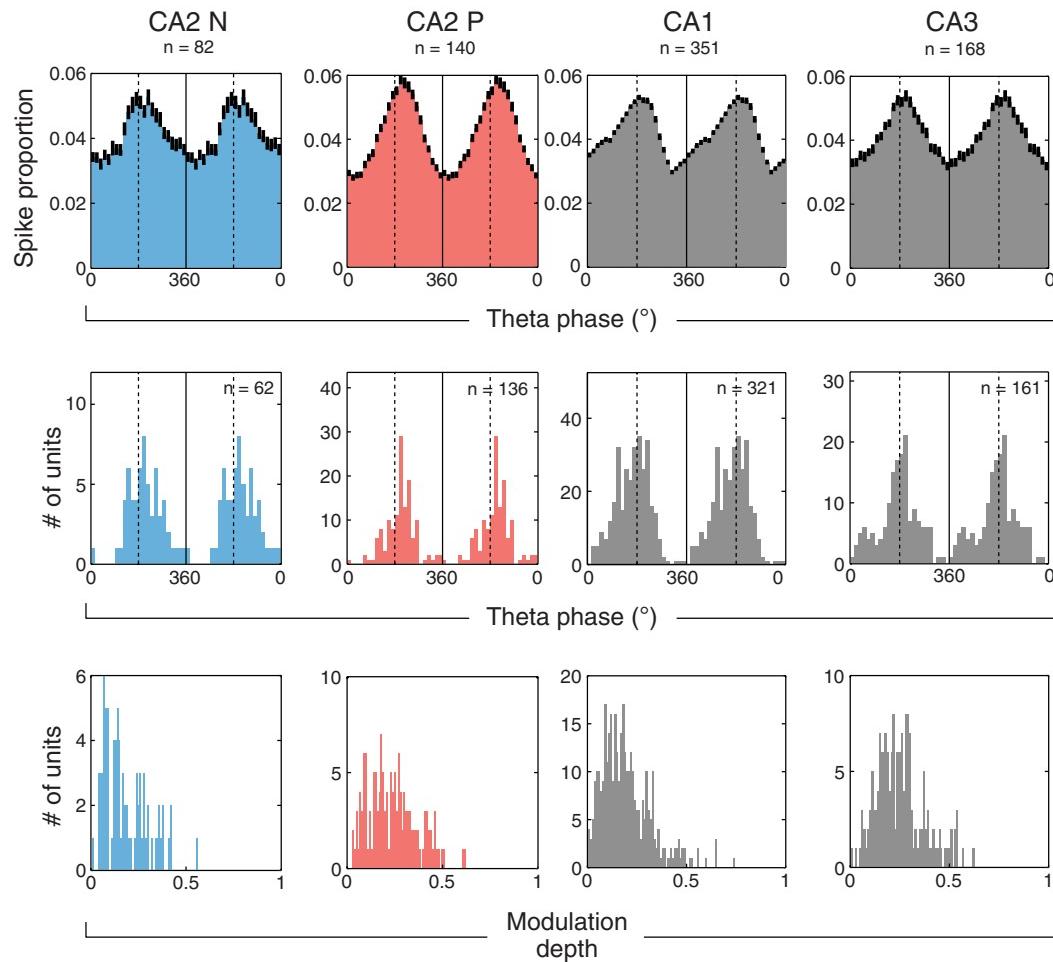
n.s., not significant at $P < 0.05$. **c**, Example spatial firing maps for CA1, CA3, CA2 P, and CA2 N units. Each column corresponds to data from an individual unit from a single 15-min task epoch. Upper row: raw maps showing positions visited by the subject (grey) and positions where the unit fired (coloured opaque points, plotted chronologically and with darker colour values at lower speeds). The total number of spikes (outside of SWR periods) in the epoch is reported at upper right. Lower two rows: occupancy-normalized firing maps, with the first row showing maps generated from data from outbound trajectories (centre to left or right arms) and the second row inbound trajectories (left or right to centre arm; Extended Data Fig. 1a). The spatial peak firing rate (highest rate for a occupancy-normalized bin) is shown at upper right. Shown are data from each unit's highest mean firing rate task epoch. Data from SWR periods were excluded from all plots. Notably, N units could show substantial firing at locations distinct from the reward wells (N unit examples with spike counts of 534, 497, 957, 1819, 668, 1,016 and 372).



Extended Data Figure 5 | See next page for caption.

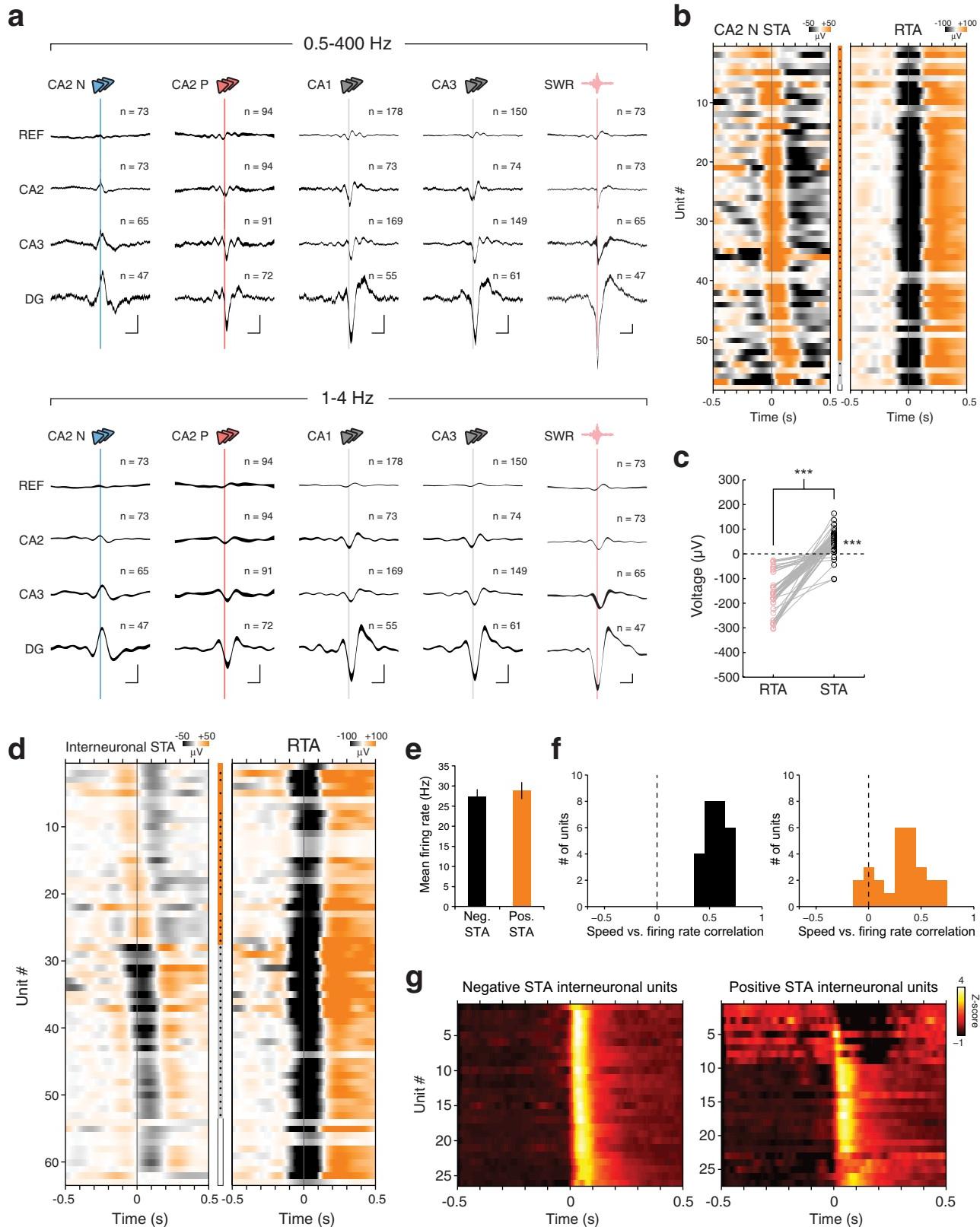
Extended Data Figure 5 | N unit spatial coding. **a**, Reward well firing rasters of 20 example N units. For each unit, data from the final ten (if available) entries of the subject's head into each of the three task reward wells (A, B, C) from a single task epoch are shown. The time of well entry ($t=0$) is plotted as a grey line. SWR periods are plotted in the background as pink zones. Note that firing for a given N unit was typically specific to one of the three reward wells. **b**, Non-reward well firing in three example N units. The rightmost example is the same as the third example in Fig. 2a. Upper row: spatial firing maps. Locations visited by the subject are plotted in grey, while locations at which the unit fired are plotted as coloured opaque points (in blue) plotted chronologically and with darker colour values at lower speeds. Total spike counts are indicated at upper right. In the task (Methods and Extended Data Fig. 1a), reward was delivered to the subjects only at the ends of the maze arms, thus locations elsewhere in the maze were not directly associated with reward. Lower row: firing rate versus speed of distinct visits to specific maze junctions (indicated with a square on spatial firing maps). Junction visits were identified as periods during which the subject's linear position (Methods) was within 10 cm of a maze junction. Firing rate was the total number of spikes divided by the visit duration. Mean speed was the average instantaneous head speed during the visit. To limit analysis to discrete traversals through a junction, visits that were both less than 1 s in duration and also had mean speeds $<10 \text{ cm s}^{-1}$ were disregarded. Note that N units tended to fire at lower speed junction visits, and that some junction visits at higher speeds

elicited no firing. **c**, Firing rate dependence on speed at non-reward task locations. Distribution of correlations (Pearson's r) between firing rate and log speed for each unit population. This analysis is the same as in Fig. 2b, except restricted to periods when the subject was located $>30 \text{ cm}$ from reward wells, moreover including only units that fired at least 50 spikes at these locations (outside of SWR periods). As in the location-inclusive case (Fig. 2b), the N unit population uniquely showed an anti-correlation ($r < 0$) of firing rate with speed. Pearson's r , mean \pm s.d.; CA1: 0.12 ± 0.20 , CA1 versus 0, $P < 10^{-23}$, signed-rank; CA3: 0.11 ± 0.18 , CA3 versus 0, $P < 10^{-13}$, signed-rank; CA2 P: 0.12 ± 0.16 , CA2 P versus 0, $P < 10^{-10}$, signed-rank; CA2 N: -0.09 ± 0.20 , CA2 N versus 0, $P = 0.0056$, signed-rank; CA2 N versus CA2 P, $P < 10^{-8}$, rank-sum. Only units with significant correlations ($P < 0.05$) were included (CA1: 386/393 units, CA3: 195/196 units, CA2 P: 121/121 units, CA2 N: 42/42 units). ** $P < 0.01$; *** $P < 0.001$. **d**, Same analysis as **c**, except with an additional restriction to periods when the subject was located in linear positions where a unit had occupancy-normalized spatial coverage $>2 \text{ Hz}$. Pearson's r , mean \pm s.d.; CA1: 0.14 ± 0.30 , CA1 versus 0, $P < 10^{-16}$, signed-rank; CA3: 0.17 ± 0.30 , CA3 versus 0, $P < 10^{-10}$, signed-rank; CA2 P: 0.22 ± 0.23 , CA2 P versus 0, $P < 10^{-12}$, signed-rank; CA2 N: -0.17 ± 0.33 , CA2 N versus 0, $P = 0.031$, signed-rank; CA2 N versus CA2 P, $P < 10^{-6}$, rank-sum. Only units with significant correlations ($P < 0.05$) were included (CA1: 358/364 units, CA3: 168/168 units, CA2 P: 111/111 units, CA2 N: 23/24 units). * $P < 0.05$; *** $P < 0.001$.

a**b**

Extended Data Figure 6 | Locomotor STAs and theta analysis. Unit spiking at speeds $>4\text{ cm s}^{-1}$ was analysed. **a**, Locomotor STAs. Plotted are mean STAs of hippocampal LFP for each principal unit population. LFP from four distinct recording sites (REF, CA2, CA3, DG) are plotted in rows. Vertical lines correspond to the time of spiking. The width of the trace indicates $\pm \text{s.e.m.}$ across individual unit STAs. The total trace length is 2 s. REF: reference electrode located in corpus callosum overlying dorsal hippocampus, reporting signals relative to a cerebellar ground screw. Scale bars, x, 250 ms; y, 50 μV . **b**, Theta phase locking analysis of each principal unit population. For comparison of theta phase preferences between unit populations in simultaneously recorded data, analysis was

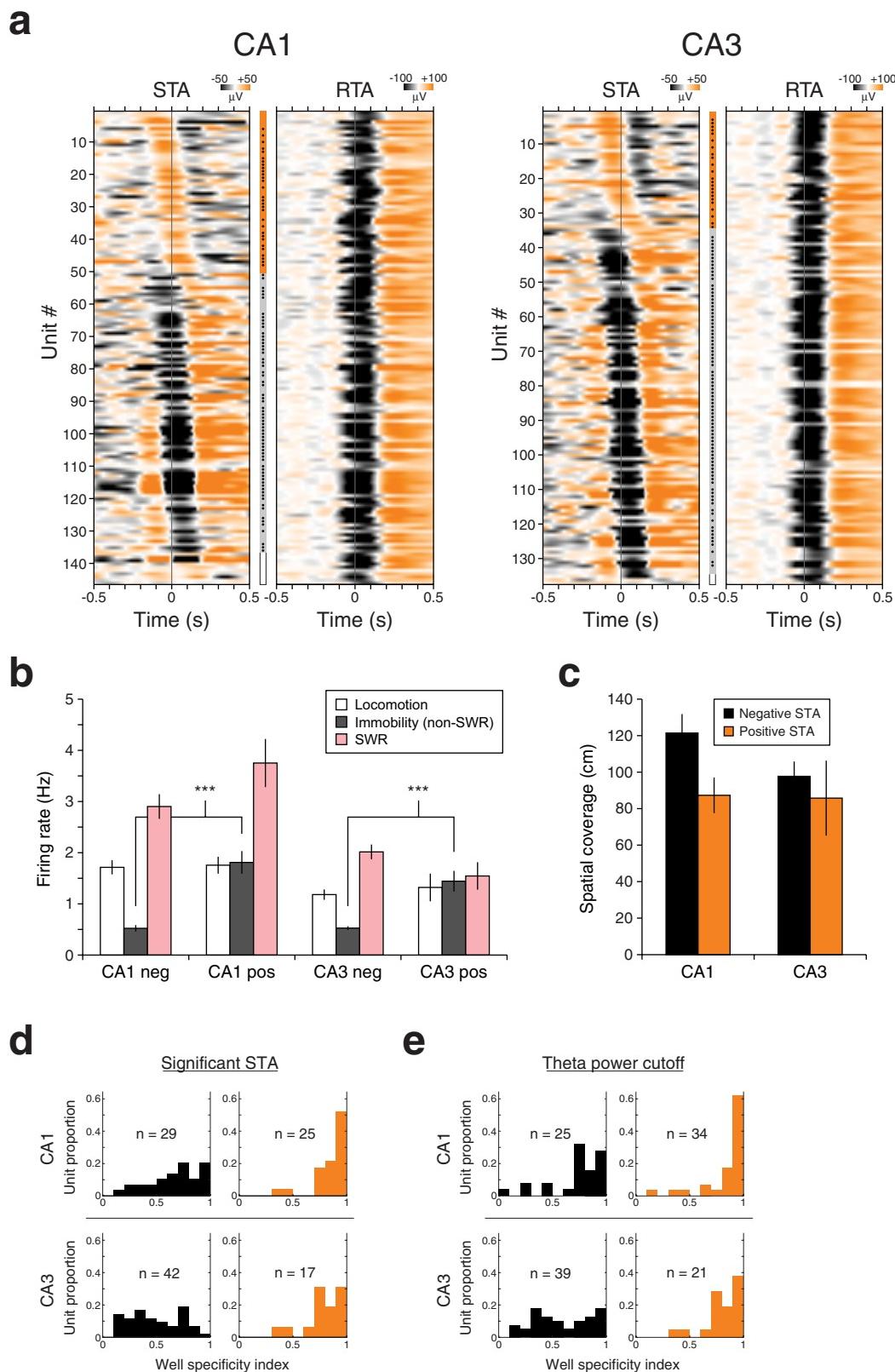
restricted to subjects in which all four unit types (CA1, CA3, CA2 N and CA2 P) were recorded. First row: mean circular distribution of spikes for each unit population. Error bars: $\pm \text{s.e.m.}$ across individual units. Second row: the distribution of mean circular phases for significantly modulated units ($P < 0.05$, Rayleigh tests, total number of significant units reported at upper right). Bottom row: the distribution of modulation depths (resultant length) for all units. In plots with theta phase (bin size: 15°; troughs at 180°, indicated in dotted lines), two cycles are shown to aid visual comparison. Surprisingly, we did not observe a $\sim 90^\circ$ phase lead of CA3 relative to CA1 as reported in a previous study³¹, perhaps due to differences in CA3 recording locations.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | N wave: a novel hippocampal network pattern at 1–4 Hz. **a**, Non-SWR immobility STAs of wide-band (0.5–400 Hz, upper section) and low frequency-band (1–4 Hz, lower section) filtered LFP. Plotted are mean STAs of hippocampal LFP for each principal unit population (first four columns). LFP from four distinct recording sites (REF, CA2, CA3, DG) are plotted in rows. The mean RTA (fifth column) was calculated from individual RTAs that were matched (same recording epochs) to each CA2 N unit, and thus have the same sample sizes as N units. Vertical lines correspond to the time of spiking (STAs) or SWRs (RTA). The width of the trace indicates \pm s.e.m. over individual unit STAs or RTAs. The total trace length is 2 s. REF: reference electrode located in corpus callosum overlying dorsal hippocampus, reporting signals relative to a cerebellar ground screw. Scale bars, x , 250 ms; y , 50 μ V. **b**, All CA2 N unit STAs for spiking during non-SWR immobility. Unit STAs are grouped by polarity at the time of spiking ($t = 0$) and sorted by the time of the extremum (peak for positive; trough for negative) nearest the time of spiking. For each unit, LFP (1–4 Hz) from CA2, CA3, or DG (in increasing order of preference when available) was used. Colours indicate voltage (colour bar). STAs are plotted on the left, while RTAs are plotted on the right. The centre bar indicates the voltage polarity of the STA (orange: positive, black: negative) at the time of spiking (STAs) or SWRs (RTAs), with a dot indicating significance versus 0 μ V ($P < 0.05$, rank-sum). The STA of an unclassified unit (see Methods) is indicated with an empty box. **c**, STA versus matched RTA voltage amplitudes (1–4 Hz LFP measured at $t = 0$; STA: time of spike, RTA: time of peak ripple power) for individual CA2 N units ($n = 58$). CA2 N unit STA amplitudes (black circles) were larger than that of their matched RTAs (pink circles) (mean \pm s.e.m., STA: $47 \pm 6 \mu$ V, RTA: $-168 \pm 10 \mu$ V; $P < 10^{-10}$, signed-rank) and also 0 μ V ($P < 10^{-7}$, signed-rank). *** $P < 0.001$. **d**, All interneuronal unit STAs

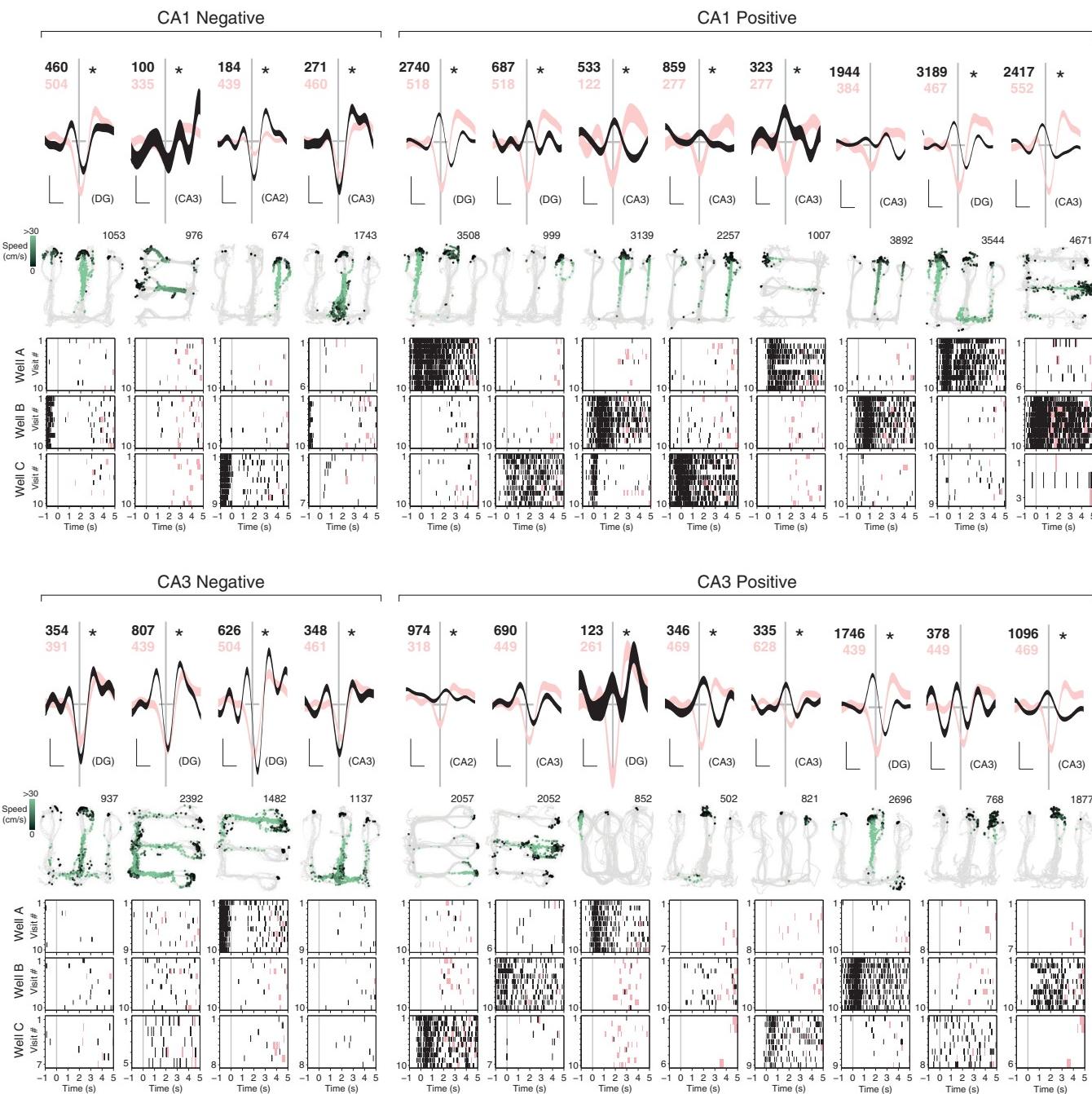
for spiking during non-SWR immobility periods. Interneuronal units were analysed for coupling to LFP since hippocampal interneurons show temporally precise firing relationships with all canonical hippocampal network patterns⁸⁰. Seventy-eight putative interneuronal units were recorded in or near the cell layers of CA1, CA2, CA3, and DG; of these units, 63 were recorded when valid CA2, CA3, or DG LFP recordings were simultaneously available and reporting SWR sharp waves as negative transients. Of the 63 units, 27 fired in association with the N wave (criteria in Methods; CA1: 10, CA2: 4, CA3: 7, and DG: 6). In the plot, unit STAs are grouped by polarity at the time of spiking ($t = 0$) and sorted by the time of the extremum (peak for positive; trough for negative) nearest the time of spiking. For each unit, LFP (1–4 Hz) from CA2, CA3, or DG (in increasing order of preference when available) was used. Colours indicate voltage (colour bar). STAs are plotted on the left, while RTAs are plotted on the right. The centre bar indicates the voltage polarity of the STA (orange: positive, black: negative) at the time of spiking ($t = 0$), with a dot indicating significance versus 0 μ V ($P < 0.05$, signed-rank). Unit STAs left unclassified (see Methods) are indicated with an empty box. **e**, Mean firing rate of interneuronal units (mean \pm s.e.m.) with negative (black; $n = 36$) versus positive (orange; $n = 27$) STAs. **f**, Firing rate versus speed correlation (Pearson's r) of interneuronal units with negative (black) versus positive (orange) STAs. Task epochs were analysed. **g**, Peri-SWR time histograms (PSTHs) of firing for interneuronal units with negative (left) and positive (right) STAs. Negative STA units uniformly exhibited a sharp peak in firing at the time of SWRs while positive STA units showed instances in which unit firing decreased from baseline levels (unit numbers 1–4, 6 and 8) or showed an increase in firing that was less sharp (unit numbers 23–25)^{80–82}.



Extended Data Figure 8 | See next page for caption.

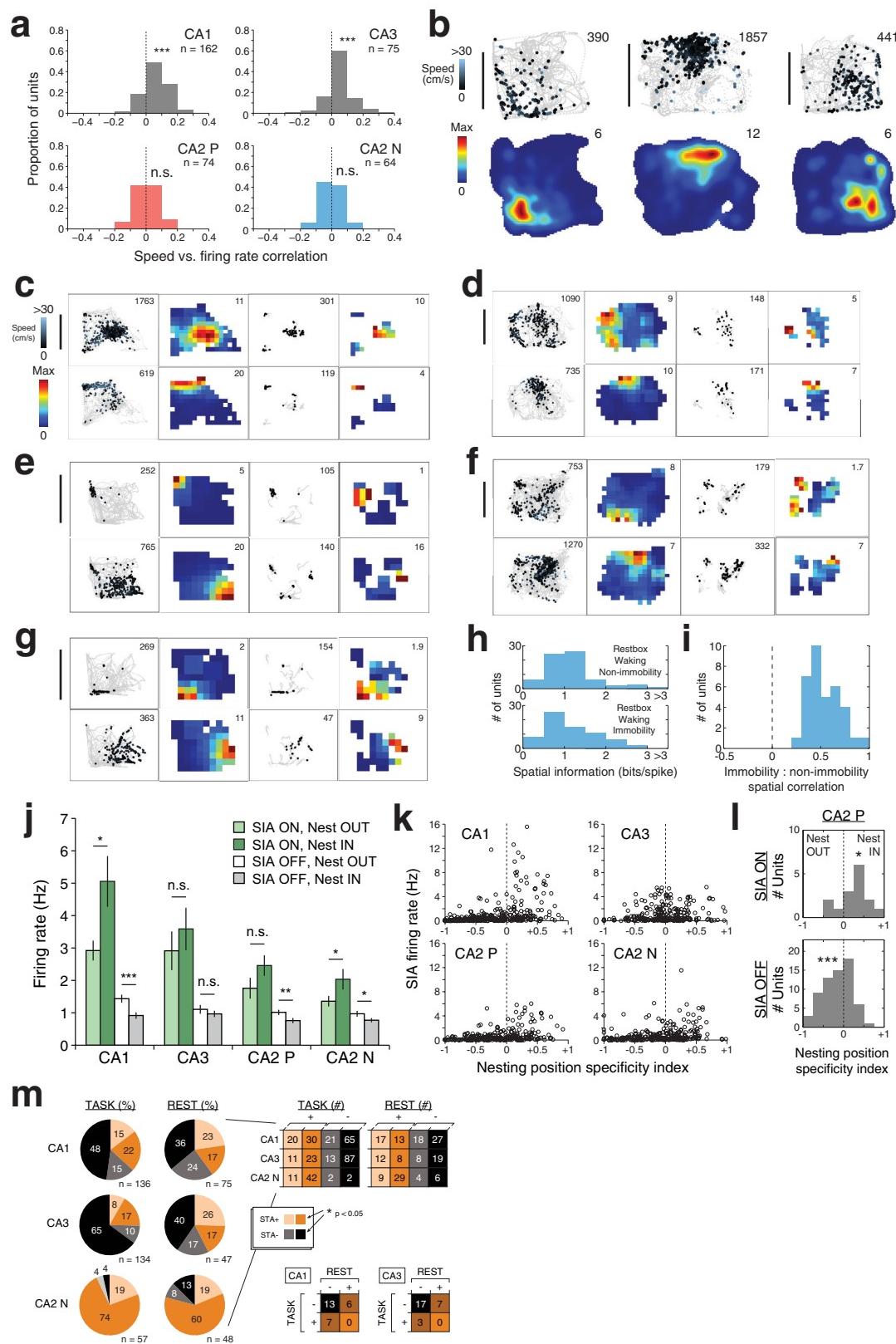
Extended Data Figure 8 | CA1 and CA3 principal neurons fire in association with the N wave. Units showing positive STAs for spiking during non-SWR immobility periods were identified as firing in association with the N wave (N wave-coupled). **a**, All CA1 and CA3 principal unit STAs for spiking during non-SWR immobility periods. Only units with >100 spikes during these periods were analysed. Unit STAs are grouped by polarity at the time of spiking ($t = 0$) and sorted by the time of the extremum (peak for positive; trough for negative) nearest the time of spiking. For each unit, LFP (1–4 Hz) from CA2, CA3, or DG (in increasing order of preference when available) was used. Colours indicate voltage (colour bar at upper right). STAs are plotted on the left, while RTAs are plotted on the right. The centre bar indicates the voltage polarity of the STA (orange: positive, black: negative) at the time of spiking ($t = 0$), with a dot indicating significance versus 0 μ V ($P < 0.05$, signed-rank). Unit STAs left unclassified (see Methods) are plotted at bottom and indicated with an empty box. **b**, Firing rates for STA-classified unit populations during task epochs (mean \pm s.e.m.; number of units: CA1 negative: 86, CA1 positive: 50, CA3 negative: 100, CA3 positive: 34). In both CA1 and CA3, units with positive STAs showed higher firing rates during non-SWR immobility (CA1 positive versus CA1 negative, $P < 10^{-9}$, rank-sum; CA3 positive versus CA3 negative, $P < 10^{-5}$, rank-sum), similar to CA2 N units (Fig. 2c). **c**, Spatial coverage in CA1 and CA3 units with negative versus positive STAs (mean \pm s.e.m.; number of units: CA1 negative: 86,

CA1 positive: 50, CA3 negative: 100, CA3 positive: 34). CA1 units with positive STAs showed somewhat lower spatial coverage than units with negative STAs (CA1 negative versus CA1 positive, $P = 0.046$, rank-sum), while an analogous difference in CA3 was not statistically significant (CA3 negative versus CA3 positive, $P = 0.12$, rank-sum). **d**, Well specificity distributions in CA1 and CA3 units that had STA amplitudes (at time of spiking) significantly different from 0 μ V (the units marked as significant in **a** and with available well data). For both CA1 and CA3, units with positive STAs showed higher well specificity (mean \pm s.e.m., CA1 negative: 0.66 ± 0.04 , CA1 positive: 0.86 ± 0.03 ; CA1 negative versus CA1 positive, $P < 10^{-4}$, rank-sum; CA3 negative: 0.49 ± 0.04 , CA3 positive: 0.79 ± 0.04 , CA3 negative versus CA3 positive, $P < 10^{-4}$, rank-sum). **e**, Well specificity distributions in CA1 and CA3 units with theta power cutoff. For each task epoch, the distribution of power in the theta band (5–11 Hz), averaged over CA1 recording sites, was calculated for immobility non-SWR periods. Spikes occurring during times in which the theta band power was in the upper quartile of this distribution were then excluded from well specificity calculations. For both CA1 and CA3, units with positive STAs showed higher well specificity (mean \pm s.e.m., CA1 negative: 0.73 ± 0.05 , CA1 positive: 0.87 ± 0.04 ; CA1 positive versus CA1 negative, $P < 0.002$, rank-sum; CA3 negative: 0.58 ± 0.04 , CA3 positive: 0.80 ± 0.04 ; CA3 negative versus CA3 positive, $P < 0.004$, rank-sum).



Extended Data Figure 9 | N wave-coupled CA1 and CA3 principal neurons. Examples of CA1 and CA3 principal units with negative versus positive STAs during non-SWR immobility. Units with positive STAs were defined as N wave-coupled. Each column corresponds to data from an individual unit. Upper sections: non-SWR immobility STA (black trace, \pm s.e.m. over individual LFP traces) and RTA (pink trace, \pm 2 s.e.m. over individual LFP traces). Vertical lines correspond to the time of spiking (for STAs) or time of SWRs (for RTAs). The total number of spikes (for STAs) and SWRs (for RTAs) averaged is reported at upper left. The region in which the LFP (at 1–4 Hz) was recorded is indicated at lower right. STAs with amplitudes (measured at the time of spiking) significantly different from 0 μ V ($P < 0.05$, rank-sum) are marked by an asterisk at

upper right. The total trace length is 1 s. A horizontal bar centred at the time of spiking indicates 0 μ V and corresponds to 200 ms. Scale bars, x, 200 ms; y, 50 μ V for STA (black trace); 100 μ V for RTA (pink trace). Middle sections: spatial firing maps. Positions visited by the subject are plotted in grey while positions at which the unit fired are shown as coloured opaque points (in green) plotted chronologically and with darker colour values at lower speeds. Shown is the 15-min task epoch in which the unit had the highest mean firing rate. The total number of spikes in the epoch is reported at upper right. Spikes occurring during SWR periods are omitted from the plots. Lower sections: well firing rasters. The time of well entry ($t = 0$) is plotted as a grey line. SWR periods are plotted in the background as pink zones.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Hippocampal spatial coding in the rest environment. **a**, Distribution of correlations (Pearson's r) between firing rate and log speed for each unit population in awake periods in the rest environment. Mean \pm s.d.; CA1 ($n = 162$ units): 0.06 ± 0.07 , CA1 versus 0, $P < 10^{-17}$, signed-rank; CA3 ($n = 75$): 0.05 ± 0.08 , CA3 versus 0, $P < 10^{-6}$, signed-rank; CA2 P ($n = 74$): 0.01 ± 0.07 , CA2 P versus 0, $P = 0.55$, signed-rank; CA2 N ($n = 64$): 0.00 ± 0.07 , CA2 N versus 0, $P = 0.77$, signed-rank, CA2 N versus CA2 P, $P = 0.47$. Only units with significant correlations ($P < 0.05$) were included (CA1: 162/163 units, CA3: 75/76, CA2 P: 74/76 units, CA2 N: 64/68 units). The N unit population did not show a significant relationship between firing rate and speed, unlike in the task environment (Fig. 2b). The positive correlation between firing rates and speed was also absent in the CA2 P population, suggesting a broader weakening of speed-dependent changes in hippocampal firing in the rest environment. This could be due to the restricted range of speeds in the rest environment enclosure and/or a fundamental influence of task conditions (Extended Data Fig. 1) on hippocampal neural activity. **b**, Three additional example N unit spatial firing maps in the rest environment. Plotted are data from awake periods. Each column corresponds to data from an individual unit. Upper row: raw maps showing positions visited by the subject (grey) and positions where the unit fired (coloured opaque points, plotted chronologically and with darker colour values at lower speeds). Total number of spikes (outside of SWR periods) in the epoch is reported at upper right. Lower row: occupancy-normalized firing maps. Peak spatial firing rate is reported at upper right. Scale bar, 20 cm. **c–g**, Awake immobility spatial firing in five example co-recorded pairs of N units from single rest recording epochs. The example pair in **c** is the same as shown at bottom in Fig. 5d. For each example pair, a unit corresponds to a row. The leftmost two columns (raw and occupancy-normalized firing maps) correspond to data from awake periods, while the rightmost two columns (raw and occupancy-normalized firing maps) correspond to data from awake immobility periods. Reported at upper right are total spike counts (raw maps) or peak spatial rates (occupancy-normalized maps). Bin size: 2.5 cm. Scale bar: 20 cm. Here, the occupancy-normalized maps shown were generated from unsmoothed occupancy-normalized maps by taking the mean firing rate of bins of a 3×3 grid centred on the bin, disregarding bins that were not occupied by the subject. Quantification in **h** and **i** was performed on unsmoothed occupancy-normalized maps. **h**, Spatial information⁸³ of N units in awake periods outside of immobility periods (upper plot, 1.12 ± 0.59 bits per spike, $n = 67$ units, with one unit excluded due to lack of firing outside of immobility) and awake immobility periods (lower plot, 1.17 ± 0.58 bits per spike, $n = 68$ units). In both conditions, data during SWR periods were excluded. Spatial information was calculated in the rest epoch in which the unit had the highest mean firing rate during awake periods. As in the task environment, N units exhibited spatially specific firing during immobility. Notably, the rest environment is an additional condition in which N units signalled location, moreover in the absence of material reward (analysis of non-reward locations in the task maze in Extended Data Fig. 5b–d). **i**, Correlation (Pearson's r) of N unit spatial maps between awake immobility periods and awake non-immobility periods in the rest environment. The correlation was calculated from unsmoothed occupancy-normalized firing maps, specifically for spatial bins in which the subject was immobile. Out of 67 units, 35 showed significant correlation ($P < 0.05$; 0.53 ± 0.03 , mean \pm s.e.m.), with no

negative correlations observed. Correlations were calculated in the rest epoch in which the unit had the highest mean firing rate during awake periods. These positive correlations indicate that N units retained their spatial specificity into immobility periods. **j**, Comparison of firing rates across SIA-nesting conditions. Statistical tests (signed-rank, comparison of Nest OUT versus IN): CA1, SIA ON ($n = 18$ units), $P = 0.014$; CA1, SIA OFF ($n = 92$), $P < 10^{-5}$; CA3, SIA ON ($n = 19$), $P = 0.60$; CA3, SIA OFF ($n = 58$), $P = 0.26$; CA2 P, SIA ON ($n = 15$), $P = 0.11$; CA2 P, SIA OFF ($n = 65$), $P = 0.0027$; CA2 N, SIA ON ($n = 18$), $P = 0.022$; CA2 N, SIA OFF ($n = 57$), $P = 0.027$. As in the evaluation of the nesting position specificity index (Fig. 5f), these comparisons show that the CA1 and CA2 N unit populations met dual criteria (description in Methods) for nesting position coding, while the CA3 unit population did not. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; n.s., not significant at $P < 0.05$. **k**, SIA firing rate versus nesting position specificity index for all detected unit-sleep period samples. Here, if data was available for a unit (in the rest unit set) during a detected sleep period, then the unit's SIA firing rate during the sleep period was measured and its nesting position specificity index was calculated with respect to that sleep period's nesting position; this sample is then represented by a scatter point. In this approach, an individual unit can contribute more than one sample. CA1 ($n = 312$ samples from 94 units): Spearman's $\rho: 0.55$, $P < 10^{-25}$. CA3 ($n = 223$ samples from 62 units): Spearman's $\rho: 0.12$, $P = 0.065$. CA2 P ($n = 263$ samples from 65 units): Spearman's $\rho: 0.37$, $P < 10^{-9}$. CA2 N ($n = 256$ samples from 60 units): Spearman's $\rho: 0.33$, $P < 10^{-7}$. **l**, CA2 P unit distribution of nesting position specificity indices. Mean \pm s.e.m.: SIA ON ($n = 15$): 0.22 ± 0.09 , $P = 0.048$, signed-rank; SIA OFF ($n = 65$): -0.16 ± 0.04 , $P < 0.001$, signed-rank. * $P < 0.05$; *** $P < 0.001$. **m**, STA class proportions across conditions. In addition to STAs calculated from non-SWR immobility in task epochs (TASK, presented in Fig. 4 and Extended Data Figs 7, 8 and 9), STAs were also calculated from non-SWR immobility during awake periods in rest epochs (REST). For REST STAs, as in TASK STAs, a minimum of 100 spikes outside of SWR periods during awake immobility and valid LFP reference sites were required, and units with STAs with mixed features were left unclassified (LFP reference site and unclassified STA criteria in Methods; unclassified unit counts: CA1: 8 out of 83, CA3: 4 out of 51, CA2 N: 10 out of 58). As in TASK, N wave-coupled units in REST were detected in substantial proportions. In left and upper right diagrams, STA positive (N wave-coupled) is in light orange, with a darker orange corresponding to significance in the STA voltage at $t = 0$ ($P < 0.05$, signed-rank). STA negative is in grey, with black corresponding to significance. Left (pie charts): proportions (%) of units in each of the STA classes. Total unit counts (number of units with classified STAs) are reported at bottom right. Percentages are rounded to nearest whole number. Upper right: unit counts in each (non-overlapping) category. Lower right: contingency table for CA1 and CA3 units found active in both task and rest epochs (fired >100 spikes outside of SWR periods during immobility in at least one task recording epoch and during awake immobility in at least one rest recording epoch) and with classifiable STAs (positive versus negative). Notably, no units were observed that were STA positive in both conditions, suggesting that N wave-coupling for a given CA1/CA3 neuron is not a static property. In contrast, the majority of classifiable CA2 N units in both TASK (53/57, or 93%) and REST (38/48, or 79%) were N wave-coupled.

Failure of RQC machinery causes protein aggregation and proteotoxic stress

Young-Jun Choe^{1*}, Sae-Hun Park^{1*}, Timm Hassemer¹, Roman Körner¹, Lisa Vincenz-Donnelly¹, Manajit Hayer-Hartl¹ & F. Ulrich Hartl¹

Translation of messenger RNAs lacking a stop codon results in the addition of a carboxy-terminal poly-lysine tract to the nascent polypeptide, causing ribosome stalling. Non-stop proteins and other stalled nascent chains are recognized by the ribosome quality control (RQC) machinery and targeted for proteasomal degradation. Failure of this process leads to neurodegeneration by unknown mechanisms. Here we show that deletion of the E3 ubiquitin ligase Ltn1p in yeast, a key RQC component, causes stalled proteins to form detergent-resistant aggregates and inclusions. Aggregation is dependent on a C-terminal alanine/threonine tail that is added to stalled polypeptides by the RQC component, Rqc2p. Formation of inclusions additionally requires the poly-lysine tract present in non-stop proteins. The aggregates sequester multiple cytosolic chaperones and thereby interfere with general protein quality control pathways. These findings can explain the proteotoxicity of ribosome-stalled polypeptides and demonstrate the essential role of the RQC in maintaining proteostasis.

Eukaryotic cells have quality control pathways to remove aberrant polypeptides from ribosomes that have stalled on mRNA, owing to mRNA truncation or the absence of a termination codon^{1–5}. Mammalian mRNAs typically contain a variable 3' untranslated region (UTR), followed by a poly(A) sequence of >60 nucleotides^{6–8}. Translation of ‘non-stop’ (NS) mRNA results in addition of a C-terminal poly-lysine tract, encoded by poly(A), which causes stalling of the NS-protein in the negatively charged ribosomal exit tunnel^{3,9,10}. The RQC complex recognizes NS-proteins and mediates their ubiquitylation and proteasomal degradation^{11–19}. The RQC comprises the E3 ubiquitin ligase Listerin (Ltn1p), Rqc1p, Rqc2p, and the AAA⁺ protein Cdc48p. Upon dissociation of the stalled ribosome^{14,20}, Rqc2p binds to the peptidyl-tRNA of the 60S subunit and recruits Ltn1p^{13,18}. The elongated Ltn1p curves around the 60S ribosome, positioning its ligase domain close to the nascent chain (NC) exit^{17–19}. Rqc2p is a nucleotide-binding protein that recruits transfer RNAs tRNA^{Ala} and tRNA^{Thr} to the 60S peptidyl-tRNA complex. This results in the addition of a C-terminal Ala/Thr sequence (CAT-tail) to the stalled NC in an mRNA-independent manner¹⁹. The CAT-tail may help clear the ribosome tunnel of stalled polypeptides.

Mutation of Listerin causes neurodegeneration in mice²¹, presumably due to a chronic defect in degrading aberrant translation products in neuronal cells. We investigated the consequences of RQC deficiency in yeast, to determine how this affects proteostasis at the molecular level, and to gain insight into the relationship between RQC and neurodegeneration.

Aggregation of non-stop proteins

To investigate the fate of NS-proteins upon RQC failure, we expressed green fluorescence protein (GFP) and firefly luciferase (Luc) from mRNAs with and without a stop codon. Only small amounts of NS-protein were detected in wild-type (WT) yeast (Fig. 1a; Extended Data Fig. 1a), consistent with efficient ubiquitylation of NS-protein (Extended Data Fig. 1b) and proteasomal degradation¹¹. In contrast, NS-protein accumulated in *ltn1Δ* cells, accompanied by the formation of SDS-resistant, high molecular weight (HMW) species (Fig. 1a). This HMW protein represented non-ubiquitylated NS-protein aggregates

(Extended Data Fig. 1b) that dissolved in formic acid (Fig. 1a). In ~17% of *ltn1Δ* cells, NS-GFP accumulated in cytosolic inclusions (Fig. 1b), independent of the RNQ prion state of the cells (Extended Data Fig. 1c, d). The vast majority of NS-GFP was not associated with ribosomes (Extended Data Fig. 1e). SDS-resistant NS-GFP aggregates were substantially smaller than ribosomes, suggesting that oligomeric aggregates coexist with visible inclusions.

Role of poly-basic tract in aggregation

The C terminus of NS-proteins contains a poly-lysine (polyK) tract encoded by the variable poly(A) tail of the mRNA. To investigate the role

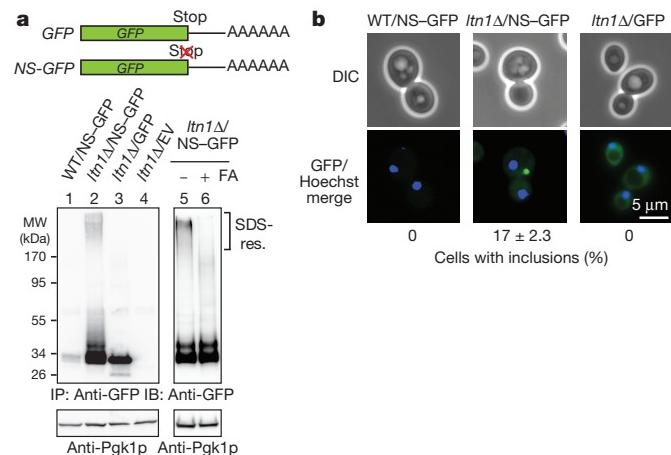


Figure 1 | NS-protein aggregates in *ltn1Δ* cells. **a**, GFP or NS-GFP was immunoprecipitated (IP) from cell extracts of WT or *ltn1Δ* yeast cells with GFP antibody, followed by anti-GFP immunoblotting (IB) (lanes 1–4). SDS-res., SDS-resistant. Pgk1p was used as loading control. EV, empty vector. NS-GFP was incubated with formic acid (FA) and analysed by IB (lanes 5, 6). **b**, Fluorescence microscopy of cells expressing GFP or NS-GFP. Nuclei stained with Hoechst 33342. *ltn1Δ* cells expressing GFP were exposed for a shorter time. The fraction of cells with visible inclusions is indicated (s.d. from 3 experiments). DIC, differential interference contrast.

¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

*These authors contributed equally to this work.

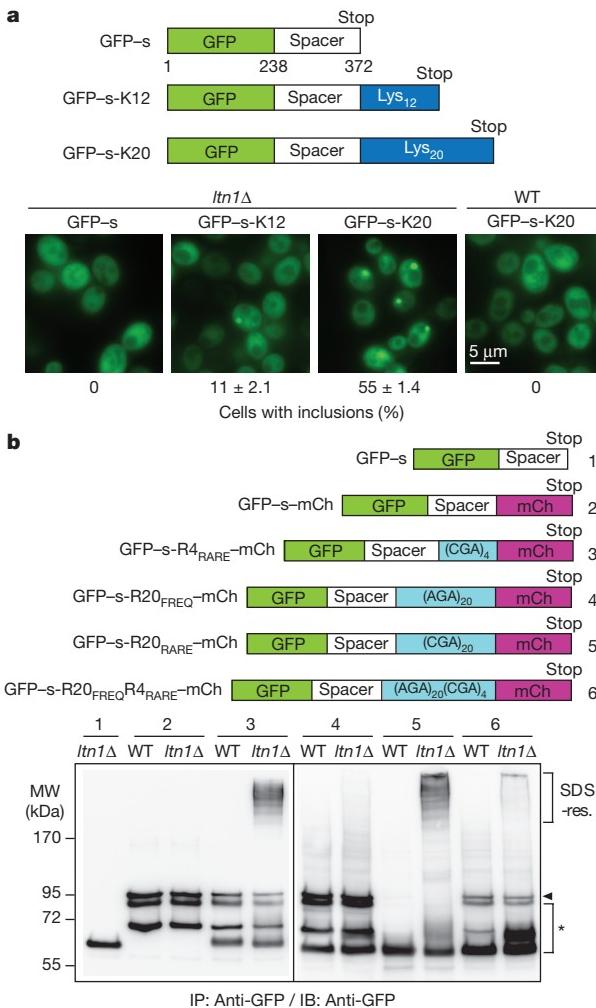


Figure 2 | Effect of poly-basic sequence and ribosomal stalling on NC aggregation. **a**, GFP fusion proteins containing an unstructured spacer (s) and poly-lysine sequence (top). Fluorescence images of *ltn1Δ* or WT cells expressing the proteins indicated (bottom). Cells with visible inclusions were quantified as in Fig. 1b. **b**, GFP-s fusion proteins containing Arg residues encoded by rare or frequent codons followed by mCherry (mCh) (top). The proteins were expressed in WT or *ltn1Δ* cells and cell extracts analysed as in Fig. 1a. Arrowhead, position of full-length protein; asterisk, stalled truncation products and proteolytic fragments.

of the C-terminal extension in aggregation, we expressed fusion proteins consisting of GFP and either a 134 amino acid spacer (GFP-s) or an additional polyK tract of 12 or 20 residues (GFP-s-K12 and GFP-s-K20, respectively) (Fig. 2a). The spacer is the unstructured, aggregation-resistant M-domain of Sup35 (refs 22, 23). GFP-s was diffusely distributed in *Ltn1Δ* cells, while GFP-s-K12 formed cytosolic inclusions in ~11% and GFP-s-K20 in ~55% of cells (Fig. 2a), independent of the spacer sequence (Extended Data Fig. 2a). Thus, aggregation depends on polyK length. Of note, the median poly(A) length in yeast is ~27 nucleotides (~10 lysines)⁸, consistent with the inclusion frequency of ~17% observed with NS-GFP (Fig. 1b). Interestingly, GFP-s-K20 formed inclusions only in *ltn1Δ* cells, but not in WT cells (Fig. 2a), suggesting that the polyK tract may mediate aggregation indirectly by causing NC stalling. These stalled chains would be degraded in WT cells.

To distinguish direct and indirect roles of the poly-basic tract in aggregation, we employed poly-arginine (polyR) stalling sequences^{12, 24} and modulated stalling efficiency by using frequent (AGA) or rare (CGA) Arg codons^{19, 25}. The GFP fusion proteins contained polyR followed by mCherry (Fig. 2b). A polyR tract of 20 residues encoded by frequent codons (GFP-s-R20_{FREQ}-mCh) efficiently produced

full-length protein (~95 kDa on SDS-PAGE) that exhibited both GFP and mCherry fluorescence, in WT and *ltn1Δ* cells (Fig. 2b; Extended Data Fig. 2b). The protein did not form SDS-resistant aggregates and was diffusely distributed, with inclusions in only ~5% of cells (Fig. 2b; Extended Data Fig. 2b). These inclusions were both GFP and mCherry fluorescent (data not shown) and thus unrelated to ribosome stalling. Stalling was observed when the polyR tract was encoded by rare codons (R4_{RARE} or R20_{RARE}), as indicated by reduced amounts of full-length protein (Fig. 2b; Extended Data Fig. 2b). These constructs produced SDS-resistant aggregates in *ltn1Δ* cells (Fig. 2b), but visible inclusions were not detectable (Extended Data Fig. 2b). Interestingly, inclusions occurred in ~23% of cells expressing a protein in which 20 Arg encoded by frequent codons are followed by 4 rare codons to mediate efficient stalling (R20_{FREQ}R4_{RARE}) (Extended Data Fig. 2b). Thus, enhanced inclusion formation requires both ribosome stalling and the translation of a long poly-basic sequence. While R20_{FREQ}R4_{RARE} formed inclusions, SDS-resistant aggregates were reduced (Fig. 2b), suggesting that the long polyR tract modulated the aggregation behaviour. Consistent with this notion, cells expressing R20_{FREQ}R4_{RARE} contained aggregates substantially larger in size than cells expressing either R4_{RARE} or R20_{RARE} (Extended Data Fig. 2c). Stalled polypeptides thus form oligomers and inclusions, of which only the former are detectable by SDS-PAGE.

In summary, RQC failure causes stalled polypeptides to accumulate in SDS-resistant aggregates. Oligomeric aggregates are already observed when the NC contains at most four arginines. NS-proteins with a poly-basic sequence exceeding a critical length of ~12 residues have an additional propensity to form visible inclusions.

Rqc2p is required for aggregation

While the function of Rqc1p is unclear, Rqc2p binds to 60S ribosomes carrying peptidyl-tRNA and recruits Ltn1p^{17–19}. Deletion of *RQC1* or *RQC2* stabilized NS-GFP, similar to deletion of *LTN1* (Fig. 3a). *Rqc1Δ* cells also accumulated SDS-resistant NS-GFP aggregates. Surprisingly, we observed no aggregates in *rqc2Δ* cells (Fig. 3a), although the NS-GFP was released from the ribosome (Extended Data Fig. 3a). Furthermore, deletion of *RQC2* in either the *ltn1Δ* or *rqc1Δ* background abolished aggregation (Fig. 3a), indicating an upstream role of Rqc2p. NS-GFP inclusions were also absent in *rqc2Δ* cells (Extended Data Fig. 3b), although other aggregation-prone proteins formed inclusions normally (Extended Data Fig. 3c). In all strains carrying the *rqc2Δ*, NS-GFP accumulated in the nucleus (Extended Data Fig. 3b), suggesting that in the absence of aggregation the polyK tract functions as a nuclear localization or retention signal. These results demonstrate that Rqc2p is required, directly or indirectly, for the aggregation of stalled NCs when the downstream degradation pathway is blocked.

CAT-tail mediates aggregation

We next tested whether aggregation is mediated by the CAT-tail added to stalled NCs by Rqc2p¹⁹. CAT-tails of 5–19 residues have been characterized by mass spectrometry¹⁹, but longer tags may exist. We observed CAT-tails with the stalled chains of the R20_{FREQ}R4_{RARE} and R20_{RARE} constructs, by comparing the band pattern of truncated chains in *ltn1Δ* and *ltn1Δrqc2Δ* cells¹⁹ (Fig. 3b). While R20_{FREQ}R4_{RARE} formed inclusions in *ltn1Δ* cells, no inclusions were detected in *ltn1Δrqc2Δ* cells (Fig. 3c), suggesting that the CAT-tail is required for the formation of both SDS-resistant aggregates and inclusions. As a critical test of this possibility we used a variant of Rqc2p, Rqc2_{aaa} (mutations D9A, D98A, R99A), which can no longer synthesize CAT-tails but recruits Ltn1p to 60S ribosomes¹⁹. As expected, *ltn1Δrqc2Δ* cells expressing WT *RQC2* added CAT-tails to stalled polypeptides, but cells expressing *rqc2aaa* did not (Extended Data Fig. 3d). However, Rqc2_{aaa} restored the ability of *rqc2Δ* cells to degrade stalled chains, reflecting recruitment of Ltn1p (Fig. 3d). Importantly, only WT Rqc2p re-established the formation of SDS-resistant aggregates and inclusions in *ltn1Δrqc2Δ*

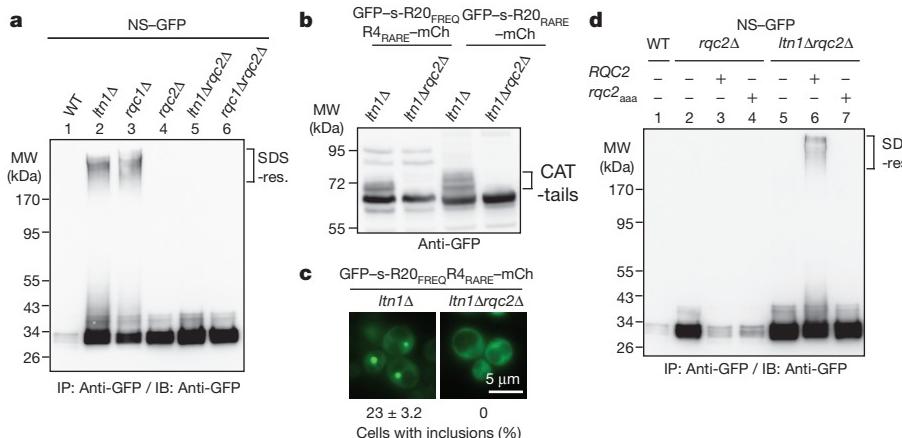


Figure 3 | Role of Rqc2p in aggregation of stalled NCs. **a**, NS-GFP was expressed in WT and RQC mutant cells, and analysed as in Fig. 1a. **b**, Extracts from *ltn1* Δ or *ltn1* Δ *rqc2* Δ cells expressing GFP-s-R20_{FREQ} R4_{RARE}-mCh or GFP-s-R20_{RARE}-mCh (see Fig. 2b) were analysed by IB with anti-GFP antibody. CAT-tails are indicated. **c**, Fluorescence images

cells (Fig. 3d; Extended Data Fig. 3e). Thus, the CAT-tail has an essential role in mediating aggregation.

To explore the role of the CAT-tail further, we generated a GFP fusion protein containing a C-terminal polyK tract of 20 residues followed by 6 Ala-Thr repeats and a stop codon (GFP-s-K20-(AT)₆) (Extended Data Fig. 4a). We expressed this protein in cells lacking Hel2p, previously implicated in ribosomal stalling¹². In this strain, stalling of the K20 tract is reduced, allowing synthesis of the full construct (Extended Data Fig. 4b). K20-(AT)₆ formed inclusions in ~54% of the cells, while proteins containing only either K20 or the (AT)₆ tag did not (Extended Data Fig. 4a). Substitution of the (Ala-Thr)₆ sequence with (Gly-Ser)₆ strongly reduced visible inclusions, confirming that both the CAT-tail and the poly-basic tract are required for inclusion formation.

NS-proteins sequester chaperones

Aggregates of neurodegenerative disease proteins often sequester molecular chaperones, impairing proteostasis^{26–28}. To identify the interactome of NS-proteins in *ltn1* Δ cells, we performed a quantitative proteomic analysis²⁹. Multiple chaperones and cofactors specifically associated with NS-GFP in *ltn1* Δ cells, prominently including the Hsp40 protein Sis1p, an essential co-chaperone of Hsp70 (ref. 30) (Fig. 4a; Supplementary Information Table 1a, b).

A substantial fraction of the NS-GFP-bound Sis1p was associated with the SDS-resistant aggregates in *ltn1* Δ and *rqc1* Δ cells (Fig. 4b; Extended Data Fig. 5a, b), and Sis1p was also recruited to NS-GFP

cells (*ltn1* Δ or *ltn1* Δ *rqc2* Δ) expressing GFP-s-R20_{FREQ} R4_{RARE}-mCh. Cells with visible inclusions were quantified as in Fig. 1b. **d**, NS-GFP was expressed in WT, *rqc2* Δ or *ltn1* Δ *rqc2* Δ cells. When indicated, the cells expressed WT Rqc2p or Rqc2_{aaa}. NS-GFP was analysed as in Fig. 1a.

inclusions (Extended Data Fig. 5c). Notably, when RQC2 was deleted, the association of Sis1p with NS-GFP was much reduced and no SDS-resistant co-aggregation occurred (Fig. 4b). Expression of WT RQC2 but not *rqc2*_{aaa} restored Sis1p co-aggregation (Extended Data Fig. 5d). Rqc2p-dependent co-aggregation of Sis1p was also observed with the stalling construct GFP-s-R4_{RARE}-mCh (Fig. 4c), which forms SDS-resistant aggregates but lacks the critical length of poly-basic sequence for inclusion formation (data not shown). These findings indicate that stalled NCs form Sis1p-associated aggregates in a CAT-tail-dependent manner, although the CAT-tail may mediate Sis1p binding indirectly.

Cells may constantly generate aberrant polypeptides that must be removed by the RQC complex^{31,32}. We found that >40% of *ltn1* Δ cells contained Sis1p-positive inclusions, even in the absence of recombinant NS-protein (Extended Data Fig. 5e). Native-PAGE of cell extracts showed that ~30% of total Sis1p was present in aggregates of ~700–1,200 kDa (Extended Data Fig. 5f). Again, expression of WT Rqc2p, but not Rqc2_{aaa}, restored the formation of Sis1p-containing aggregates in *ltn1* Δ *rqc2* Δ cells (Extended Data Fig. 5g). Thus, yeast cells accumulate considerable amounts of faulty NCs in aggregates when RQC fails, with aggregation being CAT-tail-dependent.

To assess the consequences of RQC deficiency more broadly, we analysed the spectrum of proteins associated with the SDS-resistant Sis1p aggregates by quantitative proteomics. GFP- or haemagglutinin-tagged Sis1p (expressed under the *SIS1* promoter) was immunoprecipitated

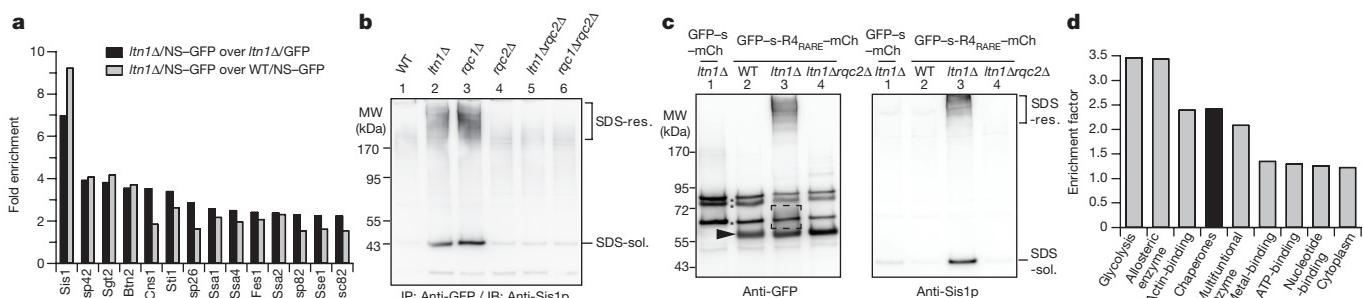


Figure 4 | Aggregates of stalled polypeptide sequester chaperones.

a, Chaperone proteins interacting with NS-GFP. Black bars, fold enrichment over GFP in *ltn1* Δ cells, and grey bars, over NS-GFP in WT cells (see Supplementary Information Table 1a, b and Methods). **b**, NS-GFP expressed in WT or RQC mutant cells was analysed by anti-GFP IP and anti-Sis1p IB. SDS-sol., SDS-soluble. **c**, GFP-s-mCh or GFP-s-R4_{RARE}-mCh (see Fig. 2b) in WT, *ltn1* Δ and *ltn1* Δ *rqc2* Δ cells.

Cell extracts were analysed by IP with anti-GFP, followed by IB with anti-GFP (left) or anti-Sis1p (right). Arrowhead, stalled truncation products; asterisk, full-length protein and proteolytic fragments; dashed box, CAT-tails. **d**, Category enrichment of proteins in SDS-resistant Sis1p aggregates (Benjamini–Hochberg false discovery rate ≤ 0.02) (see Extended Data Fig. 5h and Methods). The keyword category chaperones is highlighted in black.

from WT and *ltn1Δ* cells (Extended Data Fig. 5h). Approximately 400 proteins were reproducibly recovered in Sis1p aggregates (Supplementary Information Table 2a–c), of which ~30 were categorized as chaperones or stress proteins (Fig. 4d; Supplementary Information Table 2c), contributing ~12% to aggregate mass. Many of these chaperones were also NS–GFP interactors (~43% overlap) (Supplementary Information Tables 1a and 2c), suggesting that the recombinantly expressed NS-protein merged with the endogenous protein aggregates. Other proteins in the Sis1p aggregates are mostly localized in the cytosol and belong to various functional categories (Fig. 4d). They are typically of high abundance in the proteome³³ (Extended Data Fig. 5i), which presumably facilitated the identification of aggregated NCs. Thus, RQC deficiency causes the formation of aggregates containing numerous endogenous proteins and proteostasis components.

RQC deficiency and proteostasis stress

We next investigated whether sequestration of multiple chaperones results in proteostasis impairment of *ltn1Δ* cells. Sis1p is critical for the proteasomal degradation of terminally misfolded proteins such as cytosolic carboxypeptidase Y* (CPY*)^{27,34,35}. Indeed, CPY* fused to mCherry (CmCh*) or to GFP (CG*) was markedly stabilized in *ltn1Δ* cells compared to WT (Fig. 5a), although CPY* was efficiently polyubiquitylated (Extended Data Fig. 6a). Overexpression of Sis1p rescued degradation (Extended Data Fig. 6b). Importantly, CmCh* degradation was also restored in *ltn1Δrqc2Δ* cells (Fig. 5a), which have a normal Sis1p pool (Extended Data Fig. 5f). Thus, RQC-deficient cells fail to support general quality control pathways owing to Sis1p sequestration.

CmCh* aggregates when proteasome function is inhibited²⁷. We also observed Sis1p-positive inclusions of CmCh* in *ltn1Δ* cells (Extended Data Fig. 6c), and these co-localized with NS–GFP inclusions (Extended Data Fig. 6d), suggesting that terminally misfolded proteins and NS-proteins follow similar pathways for aggregate deposition. We note that Sis1p overexpression failed to suppress NS-protein aggregation (Extended Data Fig. 6e).

LTN1 deletion did not cause a growth defect in yeast (Extended Data Fig. 7a), despite resulting in substantial Sis1p sequestration. However, the *ltn1Δ* mutant showed slow growth upon exposure to proteostasis stress, such as CmCh* expression at 37 °C (Extended Data Fig. 7a). Overexpression of Sis1p or *RQC2* deletion rescued this growth defect (Extended Data Fig. 7b, c), consistent with CmCh* expression driving Sis1p sequestration beyond a critical level. Moreover, treatment with hygromycin B, an antibiotic that reduces translational fidelity, also caused a severe growth defect of *ltn1Δ* and *rqc1Δ* cells (Extended Data Fig. 7d), accompanied by enhanced formation of Sis1p-positive inclusions (Extended Data Fig. 7e). Again this growth defect was partially rescued by *RQC2* deletion (Extended Data Fig. 7d), suggesting that it was caused by aggregation of faulty NCs. Together these results demonstrate that RQC deficiency markedly impairs cellular proteostasis capacity.

Conclusions

Failure of ribosomal quality control, a highly evolved rescue mechanism for the removal of aberrant polypeptides, results in proteotoxic stress. We have shown that stalled nascent polypeptides aggregate when their degradation is inhibited (Fig. 5b, c). Surprisingly, the CAT-tail that is added to stalled chains by Rqc2p¹⁹ is critical in this process and is probably the major driver of aggregation of stalled polypeptides originating from truncated mRNAs³ (Fig. 5b). The aggregation of NS-proteins is more complex (Fig. 5c). In this case, read-through into the poly(A) tail of the mRNA results in the translation of a basic polyK tract that causes stalling and participates in aggregation. Our data indicate that the CAT-tail, following after the polyK tract, initiates the assembly of NS-chains to SDS-resistant oligomers, while the polyK tract mediates the formation of visible inclusions (Fig. 5c). The aggregates sequester multiple chaperones, and thereby interfere with general protein

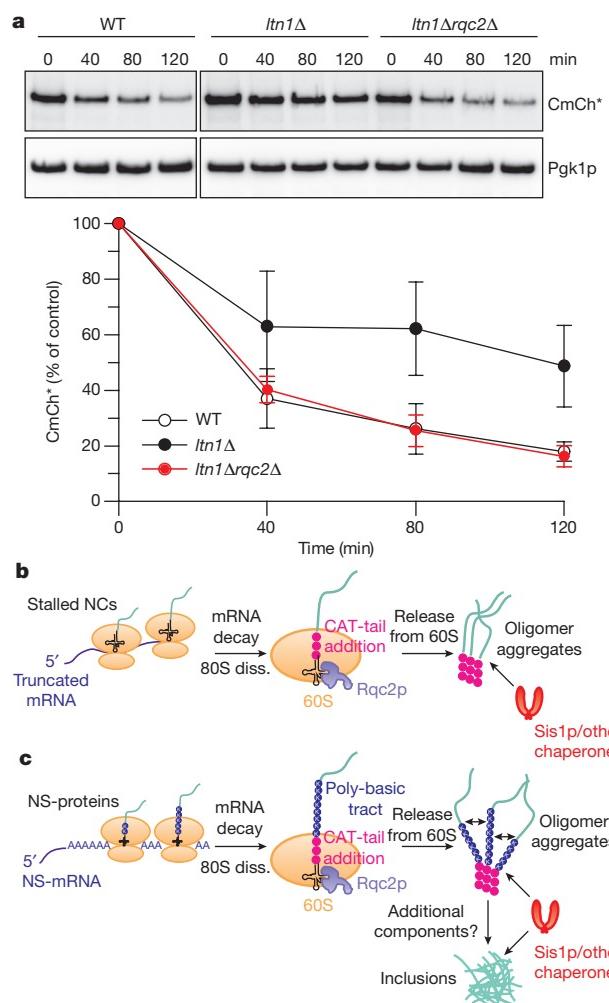


Figure 5 | Failure of RQC results in impairment of cytosolic quality control. **a**, CmCh* was expressed in WT, *ltn1Δ* and *ltn1Δrqc2Δ* cells and degradation followed by cycloheximide chase. CmCh* was detected by IB with anti-mCherry (top) and quantified by densitometry (bottom). Error bars indicate s.d. from 3 independent experiments. **b, c**, Models for the aggregation of stalled NCs (**b**) and NS-proteins (**c**), resulting in chaperone sequestration and proteostasis impairment. Stalled NCs without poly-basic tract are generated from truncated mRNAs.

quality control. The exact role of the CAT-tail in chaperone sequestration remains to be explored (Fig. 5b, c).

The SDS insolubility of the aggregates formed by stalled polypeptides suggests that the CAT-tail sequences act in a manner comparable to the poly-alanine expansions of certain disease proteins^{36,37}. The polyK tract present in NS-proteins probably contributes to aggregate formation, consistent with poly-lysine forming fibrils when charge repulsion effects are reduced at high pH³⁸. Aggregate ‘nucleation’ by the CAT-tail may serve to overcome this repulsion at physiological pH (Fig. 5c), perhaps in cooperation with negatively charged agents such as polyphosphate³⁹ or nucleic acids. Poly-basic sequences have a pronounced potential to form toxic aggregates, as exemplified by Gly-Arg or Pro-Arg dipeptide repeat sequences encoded by *C9orf72* mutant genes, which cause amyotrophic lateral sclerosis and frontotemporal dementia^{40,41}.

Defective RQC surveillance results in aggregation of a wide range of endogenous proteins and the sequestration of critical proteostasis components (Fig. 5b, c). The aggregates become highly toxic under conditions of mild conformational stress or when translational fidelity is reduced. Even in the absence of additional proteostasis pressure, the sequestration of Sis1p and other chaperones potently interferes with

cytosolic protein quality control—a positive feedback loop with the potential to cause chronic proteotoxic stress^{28,42}. Future studies will investigate whether aggregate formation by stalled polypeptides and proteostasis impairment also underlies the age-dependent neurodegenerative phenotype of the *listerin* mouse²¹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 August 2015; accepted 7 January 2016.

Published online 29 February 2016.

- Preissler, S. & Deuerling, E. Ribosome-associated chaperones as key players in proteostasis. *Trends Biochem. Sci.* **37**, 274–283 (2012).
- Pechmann, S., Willmund, F. & Frydman, J. The ribosome as a hub for protein quality control. *Mol. Cell* **49**, 411–421 (2013).
- Lykke-Andersen, J. & Bennett, E. J. Protecting the proteome: eukaryotic cotranslational quality control pathways. *J. Cell Biol.* **204**, 467–476 (2014).
- Frischmeyer, P. A. *et al.* An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258–2261 (2002).
- van Hoof, A., Frischmeyer, P. A., Dietz, H. C. & Parker, R. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**, 2262–2264 (2002).
- Barrett, L., Fletcher, S. & Wilton, S. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* **69**, 3613–3634 (2012).
- Chang, H., Lim, J., Ha, M. & Kim, V. N. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* **53**, 1044–1052 (2014).
- Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**, 66–71 (2014).
- Lu, J. & Deutsch, C. Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.* **384**, 73–86 (2008).
- Koutmou, K. S. *et al.* Ribosomes slide on lysine-encoding homopolymeric A stretches. *Elife* **4**, (2015).
- Bengtson, M. H. & Joazeiro, C. A. Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. *Nature* **467**, 470–473 (2010).
- Brandman, O. *et al.* A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* **151**, 1042–1054 (2012).
- Defenouillere, Q. *et al.* Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. *Proc. Natl Acad. Sci. USA* **110**, 5046–5051 (2013).
- Shao, S., von der Malsburg, K. & Hegde, R. S. Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation. *Mol. Cell* **50**, 637–648 (2013).
- Verma, R., Oania, R. S., Kolawa, N. J. & Deshaies, R. J. Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. *Elife* **2**, e00308 (2013).
- Shao, S. & Hegde, R. S. Reconstitution of a minimal ribosome-associated ubiquitination pathway with purified factors. *Mol. Cell* **55**, 880–890 (2014).
- Lyumkis, D. *et al.* Structural basis for translational surveillance by the large ribosomal subunit-associated protein quality control complex. *Proc. Natl Acad. Sci. USA* **111**, 15981–15986 (2014).
- Shao, S., Brown, A., Santhanam, B. & Hegde, R. S. Structure and assembly pathway of the ribosome quality control complex. *Mol. Cell* **57**, 433–444 (2015).
- Shen, P. S. *et al.* Protein synthesis. Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. *Science* **347**, 75–78 (2015).
- Shoemaker, C. J., Eyler, D. E. & Green, R. Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science* **330**, 369–372 (2010).
- Chu, J. *et al.* A mouse forward genetics screen identifies LISTERIN as an E3 ubiquitin ligase involved in neurodegeneration. *Proc. Natl Acad. Sci. USA* **106**, 2097–2103 (2009).
- Krishnan, R. & Lindquist, S. L. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* **435**, 765–772 (2005).
- Mukhopadhyay, S., Krishnan, R., Lemke, E. A., Lindquist, S. & Deniz, A. A. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc. Natl Acad. Sci. USA* **104**, 2649–2654 (2007).
- Dimitrova, L. N., Kuroha, K., Tatematsu, T. & Inada, T. Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. *J. Biol. Chem.* **284**, 10343–10352 (2009).
- Letzring, D. P., Dean, K. M. & Grayhack, E. J. Control of translation efficiency in yeast by codon-anticodon interactions. *RNA* **16**, 2516–2528 (2010).
- Olszcha, H. *et al.* Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* **144**, 67–78 (2011).
- Park, S. H. *et al.* PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell* **154**, 134–145 (2013).
- Hipp, M. S., Park, S. H. & Hartl, F. U. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* **24**, 506–514 (2014).
- Ong, S. E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols* **1**, 2650–2660 (2006).
- Yan, W. & Craig, E. A. The glycine-phenylalanine-rich region determines the specificity of the yeast Hsp40 Sis1. *Mol. Cell. Biol.* **19**, 7751–7758 (1999).
- Turner, G. C. & Varshavsky, A. Detecting and measuring cotranslational protein degradation *in vivo*. *Science* **289**, 2117–2120 (2000).
- Duttler, S., Pechmann, S. & Frydman, J. Principles of cotranslational ubiquitination and quality control at the ribosome. *Mol. Cell* **50**, 379–393 (2013).
- Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).
- Heck, J. W., Cheung, S. K. & Hampton, R. Y. Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. *Proc. Natl Acad. Sci. USA* **107**, 1106–1111 (2010).
- Summers, D. W., Wolfe, K. J., Ren, H. Y. & Cyr, D. M. The type II Hsp40 Sis1 cooperates with Hsp70 and the E3 ligase Ubr1 to promote degradation of terminally misfolded cytosolic protein. *PLoS ONE* **8**, e52099 (2013).
- Amiel, J., Trochet, D., Clement-Ziza, M., Munich, A. & Lyonnet, S. Polyalanine expansions in human. *Hum. Mol. Genet.* **13**, R235–R243 (2004).
- Forood, B., Perez-Paya, E., Houghten, R. A. & Blondelle, S. E. Formation of an extremely stable polyalanine beta-sheet macromolecule. *Biochem. Biophys. Res. Commun.* **211**, 7–13 (1995).
- Fandrich, M. & Dobson, C. M. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21**, 5682–5690 (2002).
- Gray, M. J. *et al.* Polyphosphate is a primordial chaperone. *Mol. Cell* **53**, 689–699 (2014).
- Mori, K. *et al.* The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335–1338 (2013).
- Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in C9orf72 ALS and frontotemporal dementia. *Proc. Natl Acad. Sci. USA* **110**, E4968–E4977 (2013).
- Roth, D. M. *et al.* Modulation of the maladaptive stress response to manage diseases of protein folding. *PLoS Biol.* **12**, e1001998 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Cyr, T. Inada, A. van Hoof and D. H. Wolf for reagents, and M. S. Hipp for discussions. Assistance by A. Ries and M.-J. Yoon is gratefully acknowledged. This work was supported by the European Commission under FP7 GA n°ERC-2012-SyG_318987-ToPAG, the Munich Cluster for Systems Neurology (SyNergy) and the Center for integrated Protein Science Munich (CiPSM).

Author Contributions Y.-J.C. and S.-H.P. designed and performed most of the biochemical and functional experiments. T.H. performed the sucrose gradient and semi-denaturing detergent agarose gel electrophoresis analyses. R.K. performed the mass spectrometry analysis and proteomics and L.V.-D the bioinformatics. F.U.H. and M.H.-H. supervised the experimental design and wrote the manuscript with contributions from Y.-J.C. and the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.U.H. (uhartl@biochem.mpg.de).

METHODS

Reproducibility statement. No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains. Yeast genetic experiments were carried out using standard methods. Strain BY4741 was used as the wild-type parental strain. All yeast strains used in this study are listed in Extended Data Table 1. *ltn1Δ*, *rqc1Δ* and *rqc2Δ* single deletion mutants were obtained from EUROSCARF. To delete *HEL2* and *RQC2* in the *ltn1Δ* strain, PCR-amplified marker gene expression cassettes⁴³ with overhangs complementary to upstream and downstream sequences of each gene were transformed. Addition of a C-terminal mCherry tag to *SIS1* was performed as described⁴⁴.

Plasmids. All NS-proteins, polyK and polyR expression vectors were constructed in the plasmid pRS416. The SacI-EcoRI fragment containing the *GAL1* promoter but without the *CYC1* terminator from p423GAL1⁴⁵ was ligated into pRS416. The BamHI-EcoRI fragment from pSA158 or pSA159⁴⁶ was inserted into this pRS416-*GAL1* promoter plasmid to clone the *HIS3* terminator with or without a stop codon. The PCR amplified GFP gene (including the following mutations to enhance stability and brightness: F64L, S65T, F99S, M153T, V163A, S208L⁴⁷) was inserted using XbaI-BamHI restriction sites to generate GFP and NS-GFP expression vectors. 2myc-Luc was also ligated into XbaI-BamHI sites to generate Luc and NS-Luc expression vectors. To generate the GFP-K12/K20 vectors, the *HIS3* terminator was first PCR amplified using long primers with an upstream overhang containing a (AAG)₁₂-stop codon or (AAG)₂₀-stop codon sequence. These were cloned as BamHI-EcoRI fragments. GFP was inserted using XbaI-SpeI sites and the middle domain of Sup35p (amino acid residues 124 to 253, referred to as M-domain) or a flexible region of Hsp82p (amino acid residues 210 to 263) was inserted as a natively unstructured linker^{22,23,48} using SpeI-BamHI sites.

To generate polyR vectors, GFP-SUP35M was PCR amplified without overhang sequence or with overhang sequences bearing (CGA)₄, (AGA)₂₀, (CGA)₂₀ or (AGA)₂₀(CGA)₄ sequences without a stop codon. These were ligated as XbaI-BamHI fragments. PCR amplified mCherry and the stop codon was inserted into a BamHI site.

CAT-tail constructs were generated by PCR amplifying the *HIS3* terminator using primers with (Ala-Thr)₆ or (Gly-Ser)₆ overhang sequences. The CAT-tail-*HIS3* terminator was cloned into BamHI-EcoRI sites and GFP-SUP35M or GFP-SUP35M-Lys₂₀ was cloned into XbaI-BamHI sites.

To generate *RQC2* expression constructs, the *RQC2* promoter was first cloned in p413GAL1 using SacI-BamHI sites. WT *RQC2* was PCR amplified from genomic DNA and cloned into BamHI-XhoI sites. Residues D9, D98 and R99 were mutated to alanine, resulting in *rqc2*_{aaa}¹⁹, using Q5 site-directed mutagenesis (NEB).

To generate the *Rnq1*-GFP expression construct, the *CUP1* promoter and *CYC1* terminator were cloned into pRS316 by using SacI-BamHI and XhoI-KpnI sites, respectively. *Rnq1* and GFP were cloned into BamHI-EcoRI and EcoRI-XhoI sites, respectively. The internal EcoRI site of *Rnq1* was removed by using modified PCR primers.

All plasmids used in this study are listed in Extended Data Table 2.

Immunoprecipitation of GFP and NS-GFP proteins. Cells with GFP and NS-GFP expression vector were pre-cultured in raffinose medium and then transferred to galactose/raffinose medium for ~16 h (~5 generations) at 30 °C to induce expression. Unless stated otherwise, all recombinant protein expression in this study was driven by the *GAL1* promoter under these conditions. Yeast cells were lysed with glass beads in lysis buffer A (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 5% glycerol, complete protease inhibitors; Roche) using a FastPrep-24 homogenizer with a CoolPrep adaptor (MP Biomedicals). After clearing lysates by repeated centrifugation at 2,000g for 5 min, lysates were adjusted to 2 mg ml⁻¹ protein with lysis buffer containing 0.5% NP-40. 50 µl of anti-GFP MicroBeads (Miltenyi Biotec) was added to 1 ml of final lysate. After incubation for 1 h at 4 °C, lysates with anti-GFP MicroBeads were applied to a µ column (Miltenyi Biotec). The beads were washed four times with 200 µl of lysis buffer followed by elution of bound proteins with 50 µl of HU buffer (8 M urea, 200 mM Tris-HCl pH 6.8, 1 mM EDTA, 100 mM DTT, 5% SDS, 0.01% bromophenol blue). After heating at 70 °C for 10 min, 15 µl of eluate was separated on 4–12% Bis-Tris NuPAGE gel (Invitrogen).

Immunoprecipitation of Luc and NS-Luc under denaturing condition. To preserve the ubiquitylation status of the proteins, immunoprecipitation was performed under denaturing conditions, essentially as previously reported⁴⁹. Cells were treated with 95 µM MG132 for 1.5 h before harvesting. 400 µl of 5% trichloroacetic acid (TCA) was added to a cell pellet from 20 absorbance units of cells, followed by glass bead lysis. After incubation for 1 h on ice, protein was precipitated by centrifugation and the pellet resuspended in 200 µl of 2% SDS containing 20 mM NEM (N-ethylmaleimide), 100 µM MG-132, complete protease inhibitors, and

bromophenol blue. 1 M Tris base was added until the solution turned blue. Samples were heated at 95 °C for 5 min and undissolved material was removed by centrifugation at 13,000g for 5 min. 180 µl of supernatant was diluted with 800 µl of buffer (1.2% Triton X-100, 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM EDTA, 0.5% BSA, 20 mM NEM, complete protease inhibitors). 50 µl of anti-Myc MicroBeads (Miltenyi Biotec) were added, followed by incubation for 1.5 h at 4 °C. The beads were processed as above and eluates analysed by immunoblotting with anti-Luc and anti-ubiquitin antibodies.

Polysome gradient analysis. Experiments were carried out as previously described¹¹ with minor modifications. Yeast cultures were grown to mid-log phase (*A*_{260 nm} 0.8–1.0) at 30 °C. Cycloheximide (CHX, final 0.1 mg ml⁻¹) was added 10 min before cell harvest. Cell lysates were prepared in lysis buffer B (10 mM Tris-HCl pH 7.5, 100 mM NaCl, 30 mM MgCl₂, 1 mM DTT, 0.1 mg ml⁻¹ CHX and complete protease inhibitors) using glass beads. Cell debris was removed by centrifugation at 400g for 5 min. An amount of lysate corresponding to 40 *A*_{260 nm} units was layered on a continuous 7–47% sucrose gradient prepared in 40 mM Tris-acetate pH 7.0, 50 mM NH₄Cl, 12 mM MgCl₂, 1 mM DTT and 0.1 mg ml⁻¹ CHX. Gradients were centrifuged at 40,000 r.p.m. for 2 h at 4 °C using a SW41 rotor (Beckman) and fractionated using a piston gradient fractionator coupled to an *A*_{254 nm} spectrophotometer (Biocomp). Fractions were subjected to TCA precipitation. Briefly, sodium deoxycholate was added to a final concentration of 0.02% and fractions were incubated for 15 min on ice. TCA was added to a final concentration of 10% and fractions were further incubated for 1 h on ice. Samples were then centrifuged for 30 min at 16,000g. Pellets were washed with -20 °C cold acetone and air dried. Pellets were resuspended in HU buffer and equal amounts of each fraction were loaded on a 4–12% Bis-Tris NuPAGE gel. Immunoblot analysis was carried out using anti-GFP and anti-Rpl3p antibodies.

Formic acid treatment of SDS-resistant aggregates. NS-GFP was expressed under the *GAL1* promoter in *ltn1Δ* cells and immunoprecipitated using µMACS GFP isolation kits (Miltenyi Biotec). Proteins bound to antibody beads were eluted using 100 mM triethylamine buffer (pH 11.8). After neutralizing the elutes with 1 M MES (pH 3), proteins were TCA precipitated. The pellets were washed with cold acetone and then treated with 100% formic acid at 37 °C for 1 h, followed by drying in a vacuum centrifuge concentrator. Dried proteins were re-suspended in HU buffer and heated with vigorous shaking at 65 °C for 30 min, followed by SDS-PAGE and anti-GFP or anti-Sis1p immunoblotting.

Semi-denaturing detergent agarose gel electrophoresis (SDD-AGE). SDD-AGE was performed as described previously⁵⁰. Briefly, a 1.5% agarose gel was prepared with TAE buffer (40 mM Tris base, 20 mM acetic acid, 1 mM EDTA) including 0.1% SDS. Yeast cell lysates were normalized to 4 mg ml⁻¹ and mixed with an equal volume of 2× SDD-AGE sample buffer (2× TAE, 10% glycerol, 4% SDS and 0.002% bromophenol blue) and incubated at room temperature for 10 min. 20 µg of total protein was loaded and electrophoresis was performed at 4 °C for 3 h at 75 V in TAE buffer with 0.1% SDS. After electrophoresis, proteins were transferred on a nitrocellulose membrane overnight at room temperature using the capillary transfer method with 50 mM Tris-HCl pH 7.5, 150 mM NaCl as transfer buffer, followed by immunodetection.

SILAC labelling and preparation of samples for proteomic analysis. Yeast cells were grown in synthetic complete medium with 2% raffinose without uracil and labelled with L-lysine isotopes. [¹³C₆, ¹⁵N₂]L-lysine was used as heavy lysine (H) and L-lysine D4 was used as medium lysine (M) (Cambridge Isotope Laboratories). The final concentration of lysine in the medium was 150 µg ml⁻¹. WT cells expressing NS-GFP and *ltn1Δ* cells expressing GFP were labelled with (H) and (M) lysine, respectively. *ltn1Δ* cells expressing NS-GFP were grown with normal lysine (light, L). GFP and NS-GFP expression from the *GAL1* promoter was induced by inoculating cells into the respective media containing 2% galactose and 1% raffinose. Cells were grown for at least five generations to an *A*_{260 nm} of 0.7–0.8. Immunoprecipitation of GFP and NS-GFP was carried out as described above. (H), (M) and (L) samples were mixed at 1:1:1 ratio and loaded onto 4–12% Bis-Tris NuPAGE gels. Preparation of gel slices, reduction, alkylation, and in-gel protein digestion were carried out essentially as previously reported⁵¹. Peptides were desalted, filtered, and enriched as described⁵².

NS-GFP interactome analysis by LC-MS/MS. Tryptic peptides were dissolved in 6 µl of 5% formic acid and analysed by nano LC-MS/MS using an EASY-nLC 1000 nano liquid chromatography system (Thermo) coupled to a Q-Exactive mass spectrometer (Thermo). Samples were injected onto a home-made 25 cm silica reversed-phase capillary column (New Objective) packed with 1.9-µm ReproSil-Pur C18-AQ (Dr. Maisch GmbH). Samples were loaded on the column by the nLC autosampler at a flow rate of 0.5 µl min⁻¹. No trap column was used. Peptides were separated by a stepwise 120-min gradient of 0–95% between buffer A (0.2% formic acid in water) and buffer B (0.2% formic acid in acetonitrile) at a flow rate of 250 nl min⁻¹. MS/MS analysis was performed with standard settings using cycles

of 1 high resolution (70,000 full width at half maximum (FWHM) setting) MS scan followed by MS/MS scans of the 10 most intense ions with charge states of 2 or higher at a resolution setting of 17,500 FWHM. Protein identification and SILAC based quantitation was performed with MaxQuant (version 1.3.0.5) using default settings. The UNIPROT *Saccharomyces cerevisiae* database (version 2013-12-05) was used for protein identification. MaxQuant uses a decoy version of the specified UNIPROT database to adjust the false discovery rates for proteins and peptides to below 1%.

Analysis of SDS-resistant Sis1p aggregates by LC-MS/MS. Yeast cells were grown in synthetic complete medium with 2% glucose and SILAC labelled as described above. Chromosomal *SIS1* was replaced by *SIS1*-HA or *SIS1*-GFP in WT and *ltn1* Δ cells. WT cells were isotope labelled with [$^{13}\text{C}_6$, $^{15}\text{N}_2$]L-lysine (H, heavy) and *ltn1* Δ cells were grown in normal L-lysine (L, light). Sis1p was immunoprecipitated with anti-HA or anti-GFP MicroBeads (Miltenyi Biotec). The beads from WT and *ltn1* Δ cells were eluted and the eluates mixed at a 1:1 ratio, followed by electrophoresis on 4–12% Bis-Tris NuPAGE gels. Proteins migrating above 170 kDa size were subjected to in-gel digestion and LC-MS/MS analysis (see Extended Data Fig. 4g). Proteins that were enriched ≥ 2 -fold in at least two out of three experiments each from Sis1-HA and Sis1-GFP cells (403 proteins) were defined as Sis1p aggregate interactors. The category enrichment of keywords in the set of Sis1p aggregate interactors was calculated using the Fisher exact test with a cut-off Benjamini-Hochberg false discovery rate ≤ 0.02 after annotation using Perseus (1.5.2.12). Relative abundances of proteins were estimated based on iBAQ (intensity-based absolute quantification) values (MaxQuant).

Native-PAGE analysis of cells lysates. Spheroplasts were lysed in lysis buffer C (25 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM MgCl₂, 1 mM EDTA, 5% glycerol, 0.5% Triton X-100, complete protease inhibitors) using a Dounce tissue grinder. Total lysates were centrifuged at 500g for 5 min at 4°C to remove unbroken cells. An aliquot of lysate (40 µg protein) was loaded on a 3–12% Bis-Tris native-PAGE gel (Invitrogen), followed by immunoblotting with anti-Sis1 and anti-Pgk1 antibodies. Native Protein Standard (Life Technologies) was used to estimate the molecular weight of Sis1 and its HMW form.

Cycloheximide chase. Cells grown in SC medium containing 2% glucose were transferred to medium containing 2% raffinose and 2% galactose instead of glucose. After 15–18 h of induction, CHX was added to 0.5 mg ml⁻¹ and 2.5 A_{260 nm} of cells were removed at the indicated time points. Cell extracts were prepared by alkaline lysis of cell pellets⁵³, followed by immunoblotting as above.

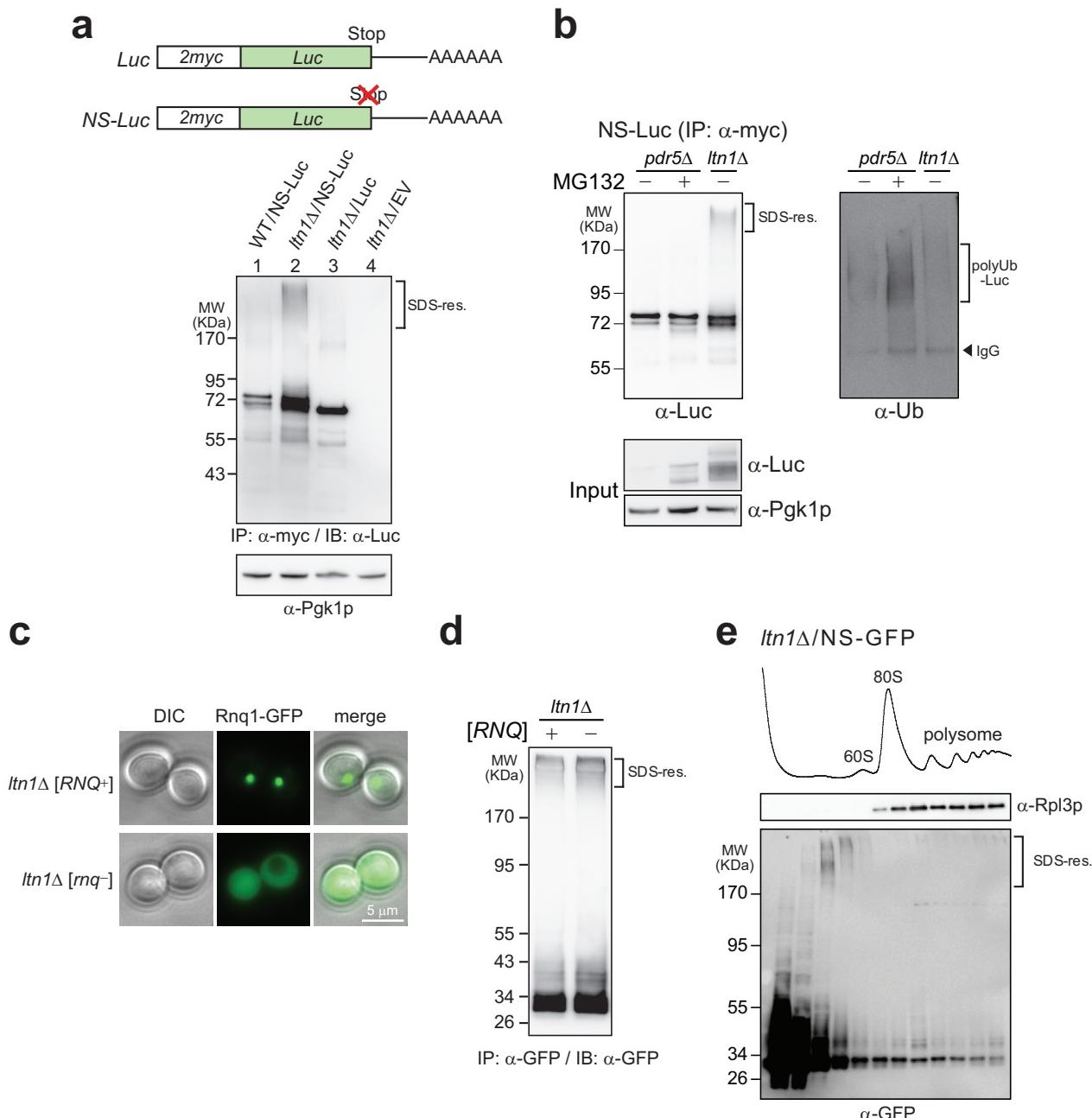
Isolation of His₆-Ub conjugated proteins. ΔssCPY* fused to mCherry (CmCh*) under the *GAL1* promoter was expressed in galactose medium for 15 h at 30°C, followed by expression of His₆-ubiquitin (His₆-Ub)⁵⁴ by addition of 100 µM CuSO₄ for 4 h. Cells were harvested and lysed with glass beads in denaturing buffer (6 M GdmCl, 100 mM NaH₂PO₄, 10 mM Tris-HCl, pH 7.0, 10 mM imidazole, 1% Triton X-100) using a FastPrep-24 homogenizer (MP biomedical). After removing cell debris (16,000g, 10 min at 4°C), lysate corresponding to 2 mg protein was incubated with 100 µl of TALON magnetic beads (Clontech) for 2 h at 4°C. Bound protein was washed three times with denaturing wash buffer (8 M urea, 100 mM NaH₂PO₄, 100 mM Tris-HCl, pH 7.8, 10 mM imidazole, 1% Triton X-100). His₆-Ub conjugated proteins were eluted with HU buffer containing 250 mM imidazole and heated for 5 min at 95°C. Eluates were separated on 4–12% Bis-Tris NuPAGE gel, followed by immunoblotting with anti-CPY antibody.

Fluorescence microscopy and image analysis. Fluorescence imaging was performed using a Zeiss Axiovert 200M inverted fluorescence microscope.

Image J and AxioVision 4.7.1 were used for image analysis. For nuclear staining, 1.5 ml of cells were collected and re-suspended in 1 ml of wash buffer (10 mM Tris-HCl pH 8.0, 10 mM MgCl₂). Cells were stained for 45 min in the dark by addition of Hoechst 33342 (final concentration 2 µg ml⁻¹) and washed three times with wash buffer before fluorescence microscopy. Cells with visible inclusions were quantified by analysing >200 cells per condition in at least three independent experiments.

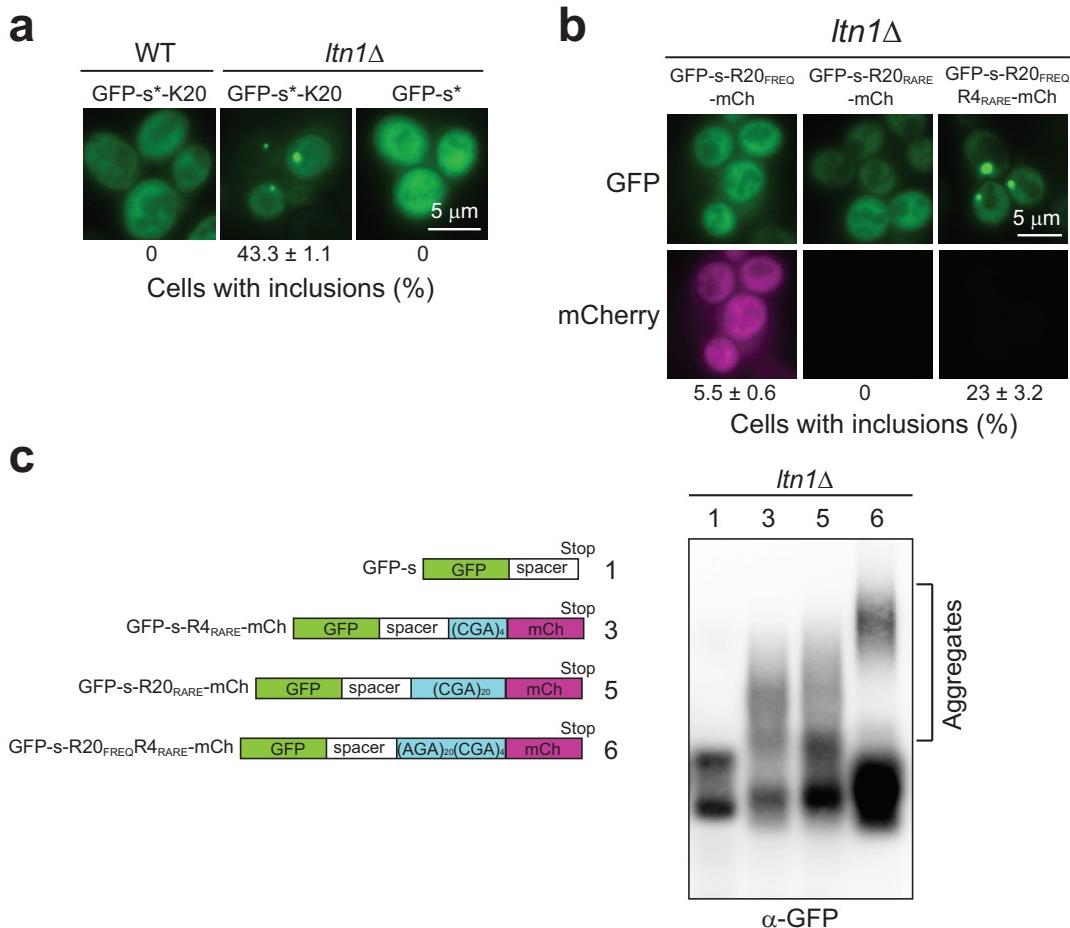
Antibodies. Anti-mCherry (Life Technologies, M11217), anti-CPY (Life Technologies, A-6428), anti-GFP (Roche, 11814460001), anti-HA (Roche, 11867423001), anti-Luciferase (Promega, G7451), anti-c-Myc (Santa Cruz Biotechnology Inc., sc-40), anti-PGK (Life Technologies, 459250), anti-Rpl3 (Developmental Studies Hybridoma Bank), anti-Sis1p (Cosmo Bio, cop-080051) and anti-ubiquitin (Santa Cruz Biotechnology Inc., sc-8017) were used for immunoblot analyses. Anti-Sis1p was a gift from D. Cyr. Anti-goat IgG-HRP (Sigma, A5420), anti-mouse IgG-HRP (Dako, P044), anti-rabbit IgG-HRP (Sigma, A6154) and anti-rat IgG-HRP (Sigma, A9037) were used as secondary antibodies for immunoblot analysis.

43. Gueldener, U., Heinisch, J., Koehler, G. J., Voss, D. & Hegemann, J. H. A second set of *loxP* marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res.* **30**, e23 (2002).
44. Young, C. L., Raden, D. L., Caplan, J. L., Czymbmek, K. J. & Robinson, A. S. Cassette series designed for live-cell imaging of proteins and high-resolution techniques in yeast. *Yeast* **29**, 119–136 (2012).
45. Mumberg, D., Muller, R. & Funk, M. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22**, 5767–5768 (1994).
46. Ito-Harashima, S., Kuroha, K., Tatematsu, T. & Inada, T. Translation of the poly(A) tail plays crucial roles in nonstop mRNA surveillance via translation repression and protein destabilization by proteasome in yeast. *Genes Dev.* **21**, 519–524 (2007).
47. Ito, Y., Suzuki, M. & Husimi, Y. A novel mutant of green fluorescent protein with enhanced sensitivity for microanalysis at 488 nm excitation. *Biochem. Biophys. Res. Commun.* **264**, 556–560 (1999).
48. Scheibel, T. et al. The charged region of Hsp90 modulates the function of the N-terminal domain. *Proc. Natl Acad. Sci. USA* **96**, 1297–1302 (1999).
49. Kragt, A., Voorn-Brouwer, T., van den Berg, M. & Distel, B. The *Saccharomyces cerevisiae* peroxisomal import receptor Pex5p is monoubiquitinated in wild type cells. *J. Biol. Chem.* **280**, 7867–7874 (2005).
50. Kryndushkin, D. S., Alexandrov, I. M., Ter-Avanesyan, M. D. & Kushnirov, V. V. Yeast [PSI⁺] prion aggregates are formed by small Sup35 polymers fragmented by Hsp104. *J. Biol. Chem.* **278**, 49636–49643 (2003).
51. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
52. Rappaport, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
53. Park, S. H. et al. The cytoplasmic Hsp70 chaperone machinery subjects misfolded and endoplasmic reticulum import-incompetent proteins to degradation via the ubiquitin-proteasome system. *Mol. Biol. Cell* **18**, 153–165 (2007).
54. Scazzari, M., Amm, I. & Wolf, D. H. Quality control of a cytoplasmic protein complex: chaperone motors and the ubiquitin-proteasome system govern the fate of orphan fatty acid synthase subunit Fas2 of yeast. *J. Biol. Chem.* **290**, 4677–4687 (2015).
55. Huh, W. K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).



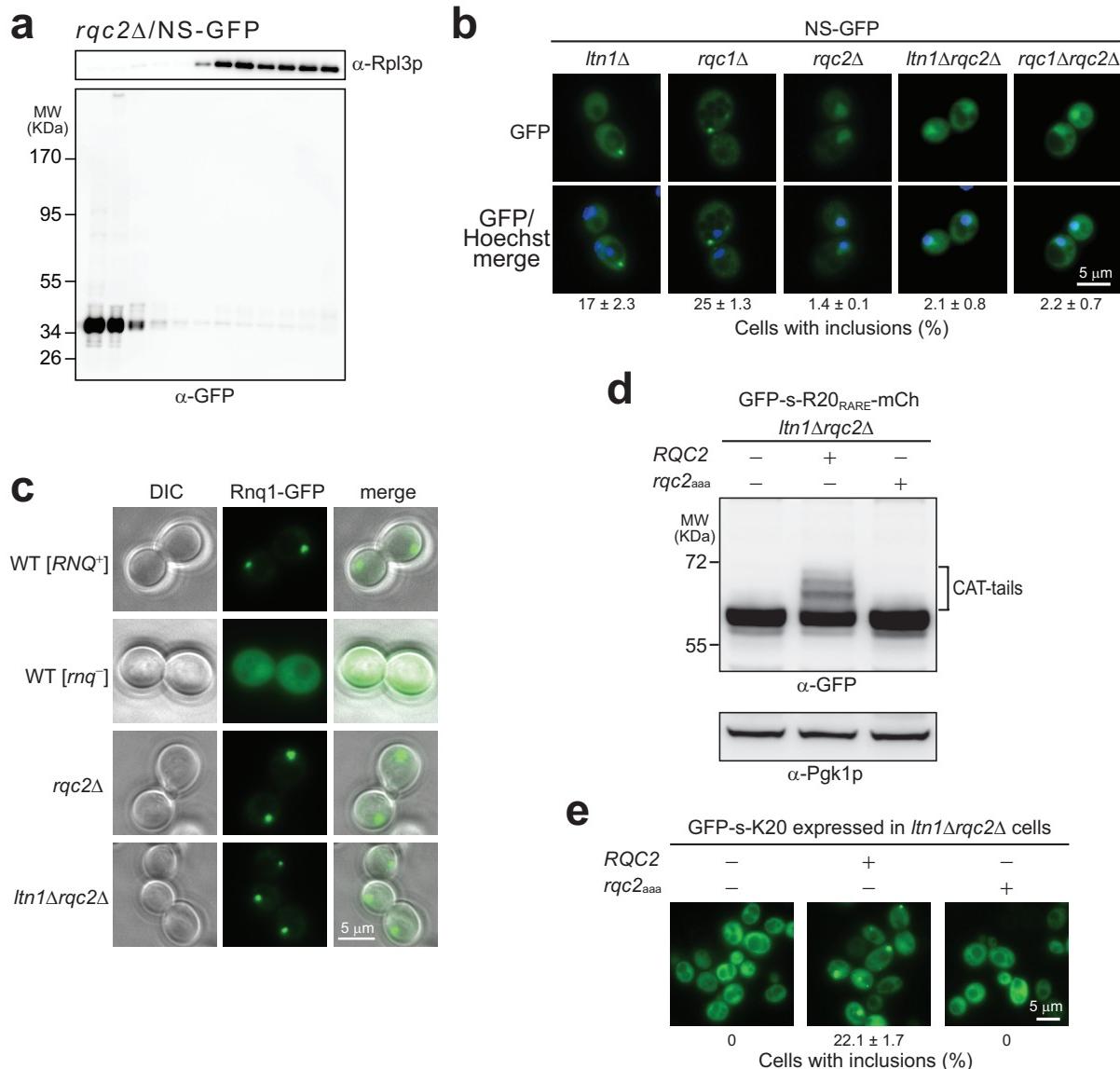
Extended Data Figure 1 | Properties of NS-proteins in *ltn1* Δ cells.
a, Firefly luciferase (Luc) or NS-Luc was expressed under the *GAL1* promoter in WT or *ltn1* Δ yeast cells for \sim 16 h (\sim 5 generations) at 30°C. Proteins were immunoprecipitated (IP) from cell extracts with anti-Myc antibody, followed by anti-Luc immunoblotting (IB). SDS-res., SDS-resistant aggregates. The cell extracts used as input were analysed by immunoblotting against phosphoglycerate kinase 1 (Pgk1p) as a loading control. EV, empty vector. **b**, SDS-resistant HMW forms of NS-Luc do not represent polyubiquitylated protein. Myc-tagged NS-Luc was expressed under the *GAL1* promoter in *pdr5* Δ or *ltn1* Δ yeast cells. *pdr5* Δ cells were incubated with DMSO or with MG132 (95 μ M in DMSO) for 1.5 h. Cell lysates were prepared under denaturing conditions (see Methods), followed by NS-Luc IP with anti-myc antibody and IB with anti-Luc antibody (left panel) or anti-Ub antibody (right panel). The positions of SDS-resistant NS-Luc, polyUb-Luc and IgG are indicated. NS-Luc and Pgk1p in input fractions were analysed. **c**, The WT yeast strain used in this study (BY4741) and its *LTN1* deletion strain were in the [RNQ $^+$] state. To cure [RNQ $^+$], cells were grown on YPD plates containing 3 mM guanidinium chloride (GdmCl) and subsequently streaked on YPD plates without GdmCl to isolate single colonies. The [RNQ $^+$] prion state was confirmed by Rnq1-GFP inclusion body formation upon expression of Rnq1-GFP from *CUP1* promoter by induction for 4 h with 50 μ M CuSO₄ during exponential growth. Live cells were analysed by fluorescence microscopy. Scale bar, 5 μ m. **d**, NS-GFP was expressed under the *GAL1* promoter in *ltn1* Δ cells in the [RNQ $^+$] or [rnq $-$] state. Cell extracts were analysed by IP and IB with anti-GFP antibody. **e**, Sucrose density gradient fractionation of *ltn1* Δ cells expressing NS-GFP for 16–18 h. Absorbance at 254 nm indicates the position of ribosomes and polysomes (top panel). Gradient fractions were immunoblotted for the 60S protein Rpl3p with anti-Rpl3p antibody (middle panel) or anti-GFP antibody (bottom panel). SDS-resistant material was incompletely recovered, presumably due to the use of 10% TCA to precipitate the fractions before IB. Note that the immunoblot was overexposed to visualize the fractionation of SDS-resistant NS-GFP.

state. To cure [RNQ $^+$], cells were grown on YPD plates containing 3 mM guanidinium chloride (GdmCl) and subsequently streaked on YPD plates without GdmCl to isolate single colonies. The [RNQ $^+$] prion state was confirmed by Rnq1-GFP inclusion body formation upon expression of Rnq1-GFP from *CUP1* promoter by induction for 4 h with 50 μ M CuSO₄ during exponential growth. Live cells were analysed by fluorescence microscopy. Scale bar, 5 μ m. **d**, NS-GFP was expressed under the *GAL1* promoter in *ltn1* Δ cells in the [RNQ $^+$] or [rnq $-$] state. Cell extracts were analysed by IP and IB with anti-GFP antibody. **e**, Sucrose density gradient fractionation of *ltn1* Δ cells expressing NS-GFP for 16–18 h. Absorbance at 254 nm indicates the position of ribosomes and polysomes (top panel). Gradient fractions were immunoblotted for the 60S protein Rpl3p with anti-Rpl3p antibody (middle panel) or anti-GFP antibody (bottom panel). SDS-resistant material was incompletely recovered, presumably due to the use of 10% TCA to precipitate the fractions before IB. Note that the immunoblot was overexposed to visualize the fractionation of SDS-resistant NS-GFP.



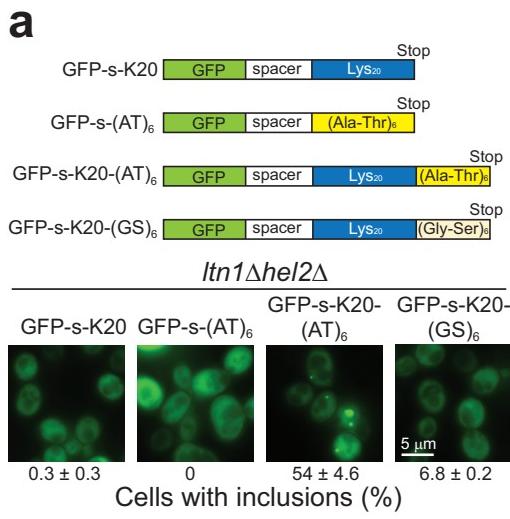
Extended Data Figure 2 | Inclusion formation by stalled poly-basic proteins in *ltn1* Δ cells. **a**, The disordered region from Hsp82p (residues 210 to 263) was used as an alternative spacer sequence (s*) in the stalling construct GFP-s*-K20, using GFP-s* as control. Representative live cell fluorescence images are shown and cells with visible inclusions were quantified as in Fig. 1b. GFP-s-K20 (Fig. 2a) and GFP-s*-K20 showed a similar frequency of inclusion formation. **b**, Live cell fluorescence microscopy of *ltn1* Δ cells expressing the GFP-s-polyR-mCh proteins indicated (see Fig. 2b). Cells were analysed for GFP and mCherry

fluorescence. Cells with visible inclusions were quantified as in Fig. 1b. Scale bar, 5 μ m. **c**, The GFP-s-polyR-mCh proteins shown on the left were expressed in *ltn1* Δ cells. Cell extracts were analysed by SDD-AGE, followed by IB with anti-GFP antibody. Note that constructs R4_{RARE} (3) and R20_{RARE} (5) form SDS-resistant aggregates detectable by SDS-PAGE (Fig. 2b), while R20_{FREQ}R4_{RARE} (6) forms inclusions but little SDS-resistant aggregates by SDS-PAGE (Fig. 2b). In lane 1, only 25% of cell lysate was applied to avoid overloading.

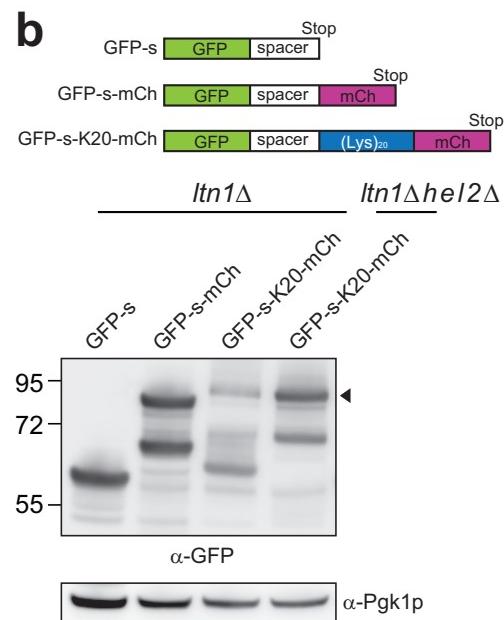


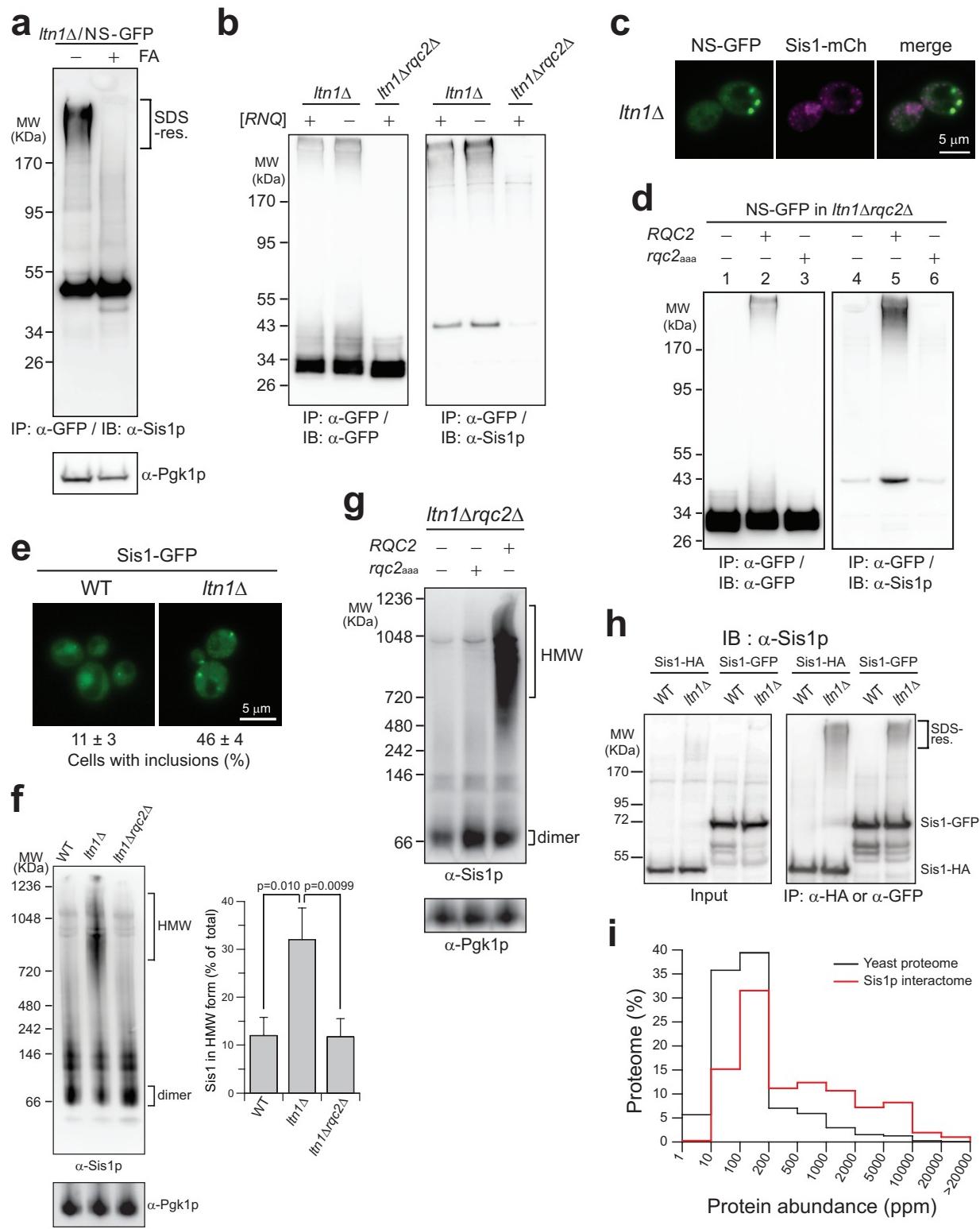
Extended Data Figure 3 | Rqc2p-dependent aggregation of stalled polypeptides. **a**, NS-GFP is released from ribosomes in *rqc2Δ* cells. Sucrose density gradient fractionation of *rqc2Δ* cells expressing NS-GFP for 16–18 h. Analysis was performed as in Extended Data Fig. 1e. **b**, Live cell fluorescence microscopy of RQC mutant cells expressing NS-GFP. Hoechst 33342 was used for nuclear staining. Cells with visible inclusions were quantified as in Fig. 1b. Scale bar, 5 μm. **c**, *rqc2Δ* cells preserve the ability to deposit aggregated protein in inclusions. The *rqc2Δ* and *ltn1Δrqc2Δ* strains used in this study were derived from the [*RNQ⁺*] WT strain. Rnq1-GFP was expressed as in Extended Data Fig. 1c to confirm inclusion formation in the *RQC2* deletion strain. WT [*rнq⁻*]

cells were isolated from WT [*RNQ⁺*] cells by GdmCl treatment as in Extended Data Fig. 1c. Inclusion formation was analysed by fluorescence microscopy. Scale bar, 5 μm. **d**, *RQC2* or *rqc2_{aaa}* was expressed under the *RQC2* promoter in *ltn1Δrqc2Δ* cells expressing GFP-s-R20_{RARE}-mCh. Cell extracts were analysed by IB with anti-GFP antibody. Pgk1p was used as a loading control. CAT-tails are indicated. **e**, *RQC2* deletion prevents inclusion formation of stalled polypeptides in *ltn1Δ* cells. GFP-s-K20 was expressed in *ltn1Δrqc2Δ* cells under the *GAL1* promoter. WT Rqc2p or Rqc2_{aaa} was co-expressed under the *RQC2* promoter in a single copy plasmid. Inclusion formation was analysed by fluorescence microscopy and quantified as in Fig. 1b. Scale bar, 5 μm.



Extended Data Figure 4 | Engineered CAT-tails mediate aggregation.
a, Schematic representation of GFP-s fusion proteins with stop codon containing 20 Lys residues or a (Ala-Thr)₆ sequence, or 20 Lys residues followed by a (Ala-Thr)₆ or (Gly-Ser)₆ sequence (top). Live cell fluorescence microscopy of *ltn1Δhel2Δ* cells expressing the proteins indicated. The fraction of cells with visible inclusions is indicated

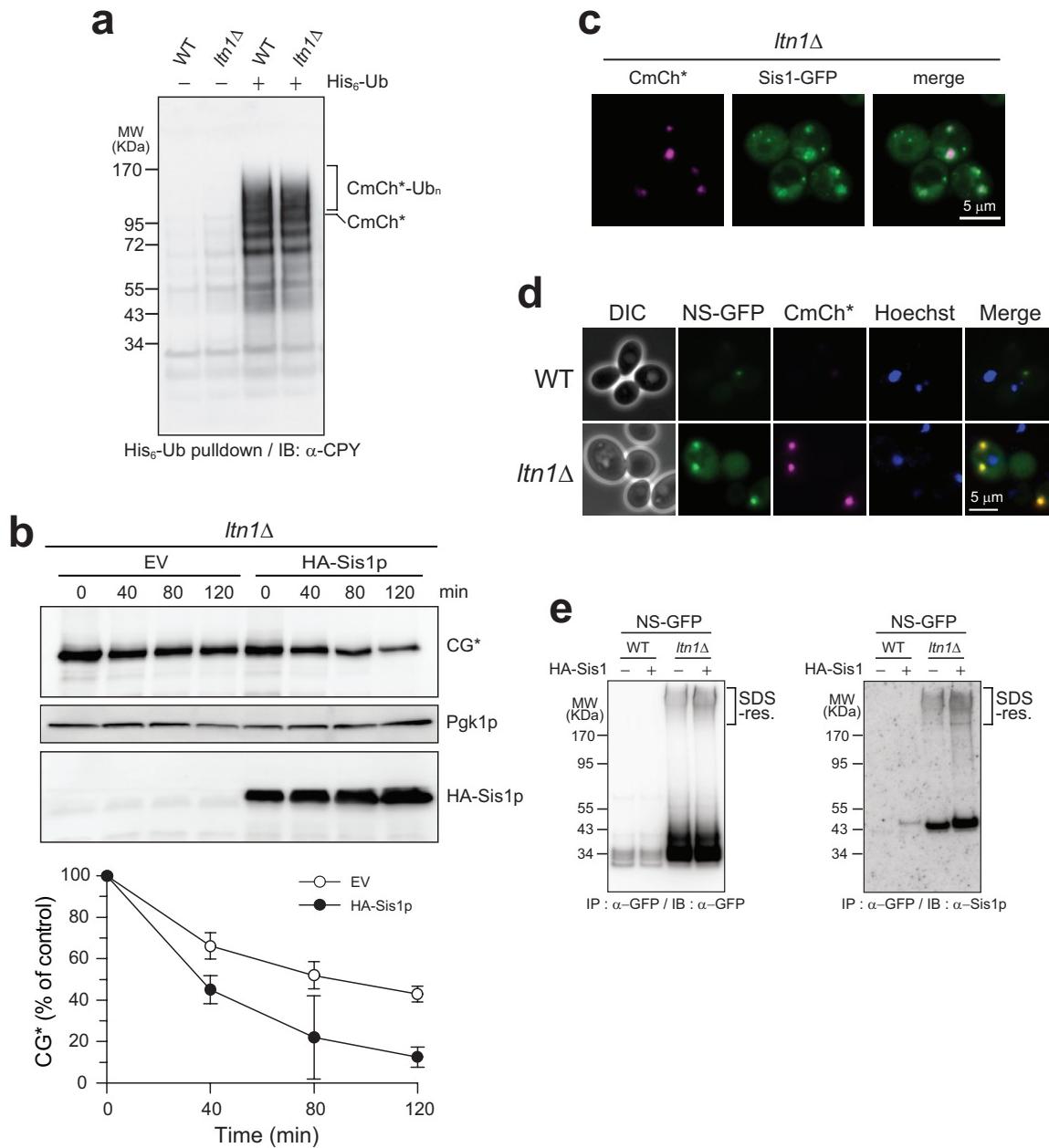




Extended Data Figure 5 | See next page for caption.

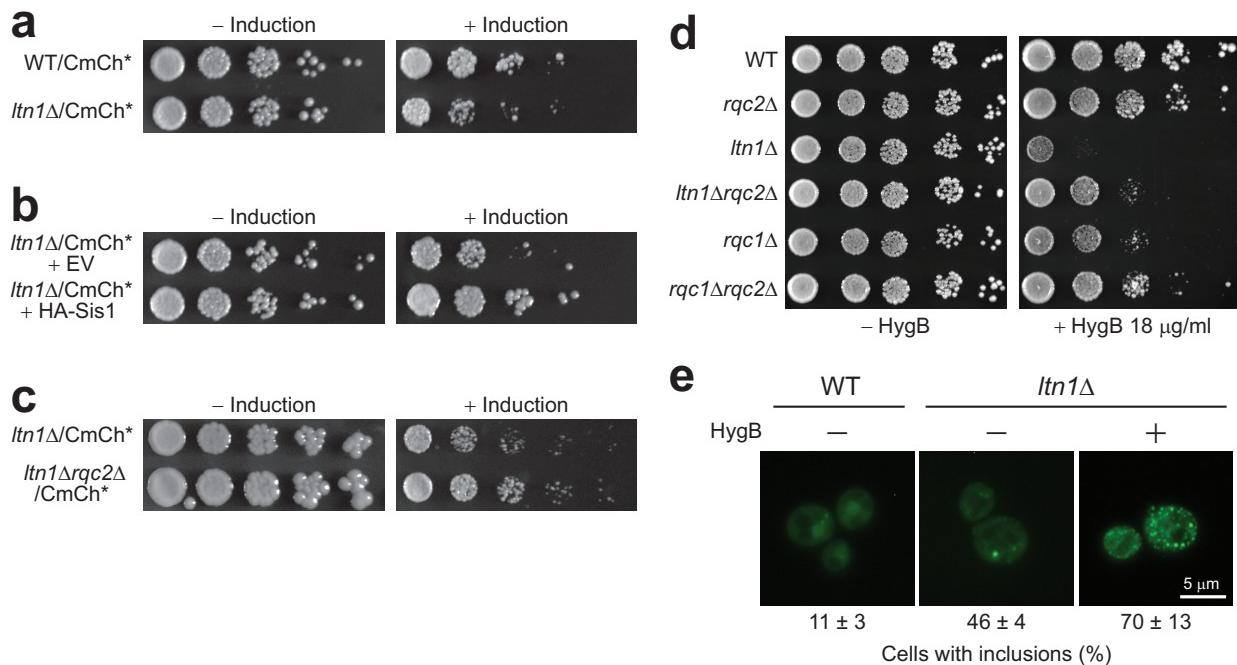
Extended Data Figure 5 | Chaperone sequestration by aggregates of stalled polypeptides. **a**, SDS-resistant co-aggregates of NS-GFP with Sis1p are solubilized with formic acid. NS-GFP was expressed in *ltn1Δ* cells and immunoprecipitated from cell extracts with anti-GFP antibody. The precipitate was incubated without or with formic acid (FA) as in Fig. 1a and analysed by IB with anti-Sis1p antibody. **b**, Formation of SDS-resistant NS-GFP aggregates and co-aggregation with Sis1p are independent of prion state. NS-GFP was expressed under the *GAL1* promoter in *ltn1Δ* [*RNQ*⁺], *ltn1Δ* [*rнq*⁻] or *ltn1Δrqc2Δ* [*RNQ*⁺] cells. Cell extracts were analysed by IP with anti-GFP, followed by IB with anti-GFP antibody (left panel) or anti-Sis1p antibody (right panel). Interaction of Sis1p and NS-GFP in [*rнq*⁻] state indicates that their interaction was not mediated by Rnq1p aggregates. **c**, Sis1-mCh co-localizes with NS-GFP inclusions in *ltn1Δ* cells. NS-GFP was expressed under the *GAL1* promoter at 30 °C in cells with *SIS1*-mCh integrated into the *SIS1* locus in the chromosome. Live cells were analysed by fluorescence microscopy. Scale bar, 5 μm. **d**, NS-GFP aggregation and co-aggregation with Sis1p in *ltn1Δrqc2Δ* cells is restored by expression of WT *RQC2* but not *rqc2_{aaa}*. NS-GFP was co-expressed with WT Rqc2p or Rqc2_{aaa} in *ltn1Δrqc2Δ* cells. NS-GFP was expressed under the *GAL1* promoter and WT Rqc2p and Rqc2_{aaa} were expressed under the *RQC2* promoter. Cell lysates were analysed by IP with anti-GFP, followed by IB with anti-GFP antibody (left panel) or anti-Sis1p antibody (right panel). **e**, Formation of Sis1p positive inclusions in *ltn1Δ* cells not expressing recombinant NS-protein. WT or *ltn1Δ* cells expressing Sis1-GFP from the genomic *SIS1* locus

were grown in YPD media at 30 °C. Cells with ≥ 2 Sis1-GFP inclusions were quantified by analysing >200 cells per condition in 4 independent experiments. Scale bar, 5 μm. **f**, Sis1p in *ltn1Δ* cells in high molecular weight (HMW) aggregates. Cell extracts from WT, *ltn1Δ* and *ltn1Δrqc2Δ* cells not expressing recombinant NS-protein were analysed by Blue native-PAGE and IB with anti-Sis1p antibody. Pgk1p was used as a loading control. The positions of the native Sis1p dimer and of HMW forms are indicated. The amount of HMW Sis1p was quantified by densitometry and expressed as percent of total. Error bars, s.d. from three independent experiments. *P* values from Student's *t*-test. **g**, Aggregation of Sis1p in *ltn1Δrqc2Δ* cells is restored by expression of WT *RQC2* but not *rqc2_{aaa}*. Extracts of *ltn1Δrqc2Δ* cells expressing WT Rqc2p or Rqc2_{aaa} under the *RQC2* promoter were analysed as in **f** without expression of recombinant NS-protein. **h**, Formation of SDS-resistant aggregates in *ltn1Δ* cells observed with Sis1-HA or Sis1-GFP. *SIS1* was chromosomally replaced by *SIS1*-HA or *SIS1*-GFP in WT or *ltn1Δ* cells. Tagged Sis1 proteins were immunoprecipitated with anti-HA antibody or anti-GFP antibody, followed by IB with anti-Sis1p antibody (right panel). Input fraction was analysed with anti-Sis1p antibody (left panel). Gel slices corresponding to the position of SDS-resistant Sis1p aggregates were excised from gels and subjected to MS-analysis to identify proteins interacting with the aggregates (see Methods). **i**, Proteins in SDS-resistant Sis1p aggregates are of relatively high abundance in the total yeast proteome. Abundance values measured in total proteome in ppm are plotted³³.



Extended Data Figure 6 | Impairment of cytosolic protein quality control in *ltn1* Δ cells. **a**, CmCh* ubiquitylation is preserved in *ltn1* Δ cells. Following expression of CmCh* from *GAL1* promoter, His₆-tagged ubiquitin (His₆-Ub) expression from *CUP1* promoter was induced with CuSO₄ for 4 h before harvesting cells. Ubiquitylated proteins were isolated by His₆ pull-down from cell lysates prepared in 6 M GdmCl to preserve polyubiquitylation. Eluates were analysed by IB with anti-CPY antibody. The positions of CmCh* and of polyubiquitylated CmCh* (CmCh*-Ub_n) are indicated. **b**, Inhibition of degradation of CPY* in *ltn1* Δ cells is rescued by overexpression of Sis1p. CPY* fused to GFP (CG*) and N-terminally HA-tagged Sis1p (HA-Sis1p) were expressed from the *GAL1* promoter. The degradation of CG* was analysed after inhibition of protein synthesis with cycloheximide as in Fig. 5a. CG* was detected by IB with anti-GFP antibody and Sis1p with anti-HA antibody (top panel). Pkg1p was used as a loading control. Data were quantified by densitometry

(bottom panel). Error bars, s.d. from three independent experiments. **c**, Sis1p co-aggregates with CmCh* in *ltn1* Δ cells. CPY*-mCherry (CmCh*) was expressed under the *GAL1* promoter at 30 °C in *ltn1* Δ cells expressing Sis1-GFP from the genomic *SIS1* locus. Live cells were analysed by fluorescence microscopy. Scale bar, 5 μ m. **d**, Live cell fluorescence microscopy of WT and *ltn1* Δ cells co-expressing NS-GFP with CmCh* for 18 h at 30 °C. Nuclei were counterstained with Hoechst 33342. DIC, differential interference contrast. Scale bar, 5 μ m. **e**, Sis1p overexpression does not suppress the formation of SDS-resistant NS-GFP aggregates in *ltn1* Δ cells. NS-GFP and N-terminally HA-tagged Sis1p (HA-Sis1p) was expressed under the *GAL1* promoter in WT or *ltn1* Δ cells. Empty vector was used as a control for HA-Sis1p. SDS-resistant aggregates were analysed by IP with anti-GFP, followed by IB with anti-GFP antibody (left panel) or anti-Sis1p antibody (right panel).



Extended Data Figure 7 | Additional proteostasis stress causes growth defect of *ltn1Δ* cells. **a–c**, Growth phenotype of RQC mutant strains. Cells from WT and RQC mutant strains indicated were transformed with CmCh* expression vector under the *GAL1* promoter and were serially fivefold diluted before spotting onto glucose medium (–Induction) and galactose/raffinose medium (+Induction). Plates were incubated for 3 days at 37 °C. In **b**, galactose inducible HA-tagged Sis1p was expressed. EV, empty vector. **d**, Hygromycin B (HygB) sensitivity of RQC mutant strains. Cells from WT and RQC mutant strains indicated were grown to exponential phase in liquid YPD medium, serially fivefold diluted

and spotted onto YPD plates with or without HygB (18 µg ml^{–1}). Plates without HygB were incubated for 2 days and with HygB for 3 days at 37 °C. **e**, Formation of Sis1p positive inclusions in *ltn1Δ* cells is enhanced under proteotoxic stress. *ltn1Δ* cells expressing Sis1–GFP replacing chromosomal *SIS1* were grown in YPD media at 30 °C without or with hygromycin B (400 µg ml^{–1}) for 18 h. Live cell fluorescence microscopy was performed. Cells with Sis1–GFP inclusions were quantified by analysing >200 cells per condition in 4 independent experiments. Data shown are an extension of the experiment shown in Extended Data Fig. 5e. Scale bar, 5 µm.

Extended Data Table 1 | Yeast strains used in this study

Name	Genotype	Source
BY4741	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	EUROSCARF
<i>pdr5Δ</i>	BY4741 <i>pdr5Δ::KanMX</i>	EUROSCARF
<i>ltn1Δ</i>	BY4741 <i>ltn1Δ::KanMX</i>	EUROSCARF
<i>rqc1Δ</i>	BY4741 <i>rqc1Δ::KanMX</i>	EUROSCARF
<i>rqc2Δ</i>	BY4741 <i>rqc2Δ::KanMX</i>	EUROSCARF
<i>ltn1Δrqc2Δ</i>	BY4741 <i>ltn1Δ::KanMX rqc2Δ::LEU2(K. lactis)</i>	This study
<i>rqc1Δrqc2Δ</i>	BY4741 <i>rqc1Δ::KanMX rqc2Δ::LEU2(K. lactis)</i>	This study
<i>ltn1Δhel2Δ</i>	BY4741 <i>ltn1Δ::KanMX hel2Δ::loxP</i>	This study
BY4741 [<i>rnl</i>]	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	This study
<i>ltn1Δ [rnl]</i>	BY4741 <i>ltn1Δ::KanMX</i>	This study
R1158	BY4741 <i>URA::CMV-tTA</i>	Open Biosystem
Tet-OFF-SIS1	R1158 <i>pSIS1::Kan-R TetO-7 TATA</i>	Open Biosystem
YPK020	BY4741 <i>sis1Δ::SIS1-mCherry-HphMX</i>	This study
YPK021	BY4741 <i>ltn1Δ::KanMX sis1Δ::SIS1-mCherry-HphMX</i>	This study
YPK022	BY4741 <i>sis1Δ::SIS1-HA-HphMX</i>	This study
YPK023	BY4741 <i>ltn1Δ::KanMX sis1Δ::SIS1-HA-HphMX</i>	This study
SIS1-GFP	BY4741 <i>sis1Δ::SIS1-GFP –HIS3</i>	Ref. 55
YPK024	SIS1-GFP <i>ltn1Δ:: HphMX</i>	This study

Ref. 55 is cited in this table.

Extended Data Table 2 | Plasmids used in this study

Name	Expression	Source
pYJC001	p416GAL1-NS Luc	This study
pYJC002	p416GAL1-Luc	This study
pYJC003	p416GAL1-NS GFP	This study
pYJC004	p416GAL1-GFP	This study
pYJC005	p416GAL1-GFP-spacer	This study
pYJC006	p416GAL1-GFP-spacer-K12	This study
pYJC007	p416GAL1-GFP-spacer-K20	This study
pYJC008	p416GAL1-GFP-spacer-mCherry	This study
pYJC009	p416GAL1-GFP-spacer-R4 _{RARE} -mCherry	This study
pYJC010	p416GAL1-GFP-spacer-R20 _{FREQ} -mCherry	This study
pYJC011	p416GAL1-GFP-spacer-R20 _{RARE} -mCherry	This study
pYJC012	p416GAL1-GFP-spacer-R20 _{FREQ} R4 _{RARE} -mCherry	This study
pYJC013	p416GAL1-GFP-spacer-(Ala-Thr)6	This study
pYJC014	p416GAL1-GFP-spacer-K20-(Ala-Thr)6	This study
pYJC015	p416GAL1-GFP-spacer-K20-(Gly-Ser)6	This study
pYJC016	p416GAL1-GFP-spacer-K20-mCherry	This study
pYJC017	p416GAL1-GFP-s* (Hsp82 flexible region)	This study
pYJC018	p416GAL1-GFP-s*-K20 (Hsp82 flexible region)	This study
pYJC019	p316CUP1-RNQ1-GFP	This study
pYJC020	p413RQC2-WT RQC2	This study
pYJC021	p413RQC2- rqc2 _{aaa}	This study
pGAL-CmCh*	p413GAL-ΔssCPY*-mCherry (CPY*-mCh)	Ref. 27
pGAL-CG*	p413GAL-ΔssCPY*-GFP	Ref. 27
pHA-SIS1	p415GAL-HA-SIS1	Ref. 27
pWO1125	YEplac181CUP-His ₆ -ubiquitin	Ref. 54
RPB42	p413GPD-ΔssPrA	Ref. 54

Structure of the voltage-gated two-pore channel TPC1 from *Arabidopsis thaliana*

Jiangtao Guo¹, Weizhong Zeng^{1,2}, Qingfeng Chen^{1,2}, Changkeun Lee^{1,2}, Liping Chen^{1,2}, Yi Yang^{1,2}, Chunlei Cang³, Dejian Ren³ & Youxing Jiang^{1,2}

Two-pore channels (TPCs) contain two copies of a *Shaker*-like six-transmembrane (6-TM) domain in each subunit and are ubiquitously expressed in both animals and plants as organellar cation channels. Here we present the crystal structure of a vacuolar two-pore channel from *Arabidopsis thaliana*, AtTPC1, which functions as a homodimer. AtTPC1 activation requires both voltage and cytosolic Ca^{2+} . Ca^{2+} binding to the cytosolic EF-hand domain triggers conformational changes coupled to the pair of pore-lining inner helices from the first 6-TM domains, whereas membrane potential only activates the second voltage-sensing domain, the conformational changes of which are coupled to the pair of inner helices from the second 6-TM domains. Luminal Ca^{2+} or Ba^{2+} can modulate voltage activation by stabilizing the second voltage-sensing domain in the resting state and shift voltage activation towards more positive potentials. Our Ba^{2+} -bound AtTPC1 structure reveals a voltage sensor in the resting state, providing hitherto unseen structural insight into the general voltage-gating mechanism among voltage-gated channels.

TPCs are cation channels ubiquitously expressed in the organelles of animals and plants^{1–4} (Extended Data Fig. 1a) and believed to be evolutionary intermediates between homotetrameric voltage-gated potassium/sodium channels and four-domain single-subunit voltage-gated sodium/calcium channels⁵. Each TPC subunit contains 12 transmembrane segments that can be divided into two homologous copies of an S1–S6 *Shaker*-like 6-TM domain⁶, with the channel assembling as a dimer—the equivalent of a voltage-gated tetrameric cation channel.

Since the molecular identification of the first TPC channel from rat kidney⁷, three subfamilies of animal TPC channels have been defined—TPC1, TPC2 and TPC3—with the first two expressed ubiquitously in animals and the subject of extensive studies^{2,8–15}. Animal TPC1 and TPC2 are localized to the endosomal/lysosomal membrane and their physiological functions are still under debate. While some studies suggested TPCs mediate nicotinic acid adenine dinucleotide phosphate (NAADP)-dependent calcium release from endolysosomes^{2,8,9,15}, others have proposed they are sodium-selective channels activated by PI(3,5)P₂ rather than NAADP^{10,11}. It has also been shown that mammalian TPCs interact with the mTOR complex and sense cellular nutrient status via ATP inhibition in an mTOR-dependent manner¹¹. A recent study demonstrated that TPC activity is essential for the release of Ebola virus from endosome/lysosome into the host cell, thus making TPCs potential targets for the treatment of Ebola infection¹⁶.

AtTPC1, the first TPC channel cloned from a plant¹⁷, is localized to the vacuolar membrane and is responsible for generating the slow vacuolar (SV) current that was observed long before the channel's molecular identification¹⁸. Consequently, AtTPC1 is also called the SV channel. AtTPC1 is a non-selective cation channel, permeable to various monovalent cations as well as Ca^{2+} (refs 19, 20) and probably has an important role in regulating cytosolic ion concentrations³. The channel is voltage-gated and its voltage-dependent activation can be modulated by both cytosolic and vacuolar Ca^{2+} . Cytosolic Ca^{2+} potentiates voltage activation by binding to the EF-hand domain, located between the two 6-TM domains in plant TPC1 but absent in animal TPCs²¹. Notably, vacuolar Ca^{2+} adversely affects channel gating by slowing down voltage

activation and shifting the voltage dependence towards positive potentials²². It has been shown that plant TPCs are involved in the regulation of various physiological processes such as germination and stomatal opening¹, jasmonate biosynthesis^{23,24}, and long-distance calcium wave propagation induced by high salt concentrations²⁵. In this study, we determined the crystal structure of AtTPC1 to 3.3 Å resolution, which, along with electrophysiological analysis, reveals the molecular mechanism of voltage-gating and calcium modulation in plant TPC1.

Functional analysis of AtTPC1

Unlike most activity measurements of AtTPC1 channels employing direct patch clamp recording of vacuolar membranes, we expressed AtTPC1 in HEK293 cells and measured plasma membrane channel activity using whole-cell patch clamping (Extended Data Fig. 2a and Methods). In this setting, the extracellular side (facing the bath solution) is equivalent to the luminal side of AtTPC1 in vacuoles. As previously shown, AtTPC1 is voltage-gated and cytosolic Ca^{2+} is required for channel activation, as no current was observed at 100 mV membrane potential at $[\text{Ca}^{2+}]_{\text{cytosolic}}$ below 100 nM (Fig. 1a). Cytosolic Ca^{2+} potentiates channel activation by shifting the voltage activation towards hyperpolarization, increasing the activation rate and slowing down deactivation. Conversely, increasing bath $[\text{Ca}^{2+}]$, analogous to increased vacuolar Ca^{2+} , shifts voltage activation towards a more positive potential, with the channel displaying slowed activation and faster deactivation (Fig. 1b). Ba^{2+} can have a similar inhibitory effect as vacuolar Ca^{2+} (Extended Data Fig. 2b). The non-selective nature of AtTPC1 was assessed using Na^+ and K^+ as permeating ions, confirming that AtTPC1 conducts Na^+ and K^+ equally well (Fig. 1c). No channel inactivation was observed in any of our recordings.

Overall structure of AtTPC1

The crystal structure of AtTPC1, determined to 3.3 Å (Methods and Extended Data Table 1) reveals two 6-TM domains (6-TM I and 6-TM II) and an intervening cytosolic EF-hand domain per AtTPC1 subunit, two of which assemble into a functional channel equivalent to a tetrameric voltage-gated channel (Fig. 2a–c). Following the same

¹Department of Physiology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9040, USA. ²Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9040, USA. ³Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

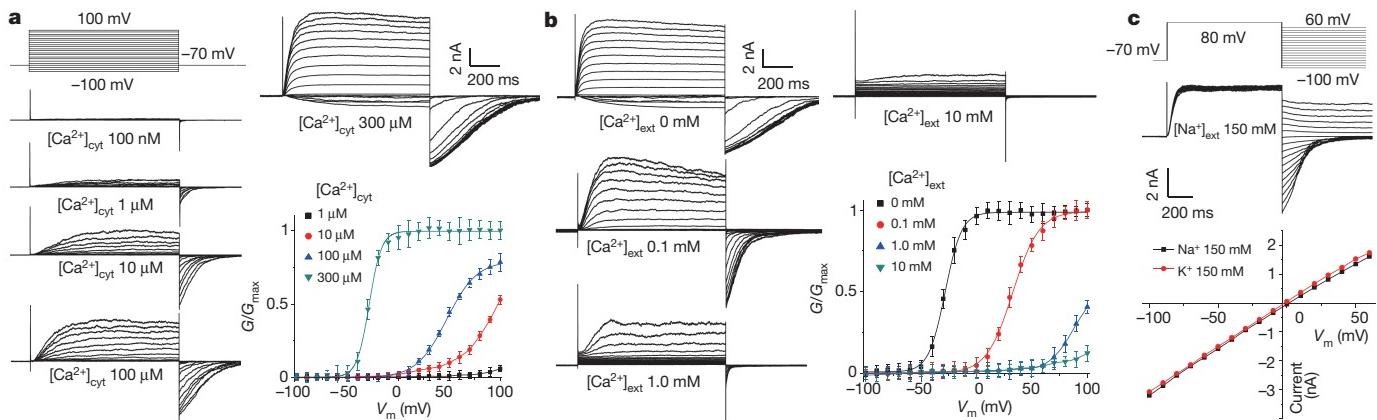


Figure 1 | Voltage activation and Ca^{2+} modulation of AtTPC1 overexpressed in HEK cell. **a**, Cytosolic Ca^{2+} -dependent voltage activation of AtTPC1. Currents were recorded with varying $[\text{Ca}^{2+}]_{\text{cytosolic}}$ in pipette and calcium-free in bath (extracellular). Boltzmann fit yields $V_{1/2} = -28 \text{ mV}$, $Z = 3.9$ ($V_{1/2}$, membrane potential for half maximum activation; Z , apparent valence) for voltage activation in $300 \mu\text{M}$ $[\text{Ca}^{2+}]_{\text{cytosolic}}$ and $V_{1/2} = 48 \text{ mV}$, $Z = 1.9$ in $100 \mu\text{M}$ $[\text{Ca}^{2+}]_{\text{cytosolic}}$. V_m , membrane potential; G , conductance. **b**, Extracellular Ca^{2+} inhibition

of AtTPC1. Currents were recorded with the presence of $300 \mu\text{M}$ $[\text{Ca}^{2+}]_{\text{cytosolic}}$ (pipette) using the same protocol as **a**. $V_{1/2} = -28 \text{ mV}$, $Z = 3.5$ for $0 \text{ } [\text{Ca}^{2+}]_{\text{extracellular}}$; $V_{1/2} = 33 \text{ mV}$, $Z = 2.1$ for 0.1 mM $[\text{Ca}^{2+}]_{\text{extracellular}}$. **c**, Selectivity measurement of AtTPC1 with intracellular 150 mM Na^+ , $300 \mu\text{M}$ Ca^{2+} , and extracellular 150 mM Na^+ or K^+ , 0 mM Ca^{2+} . Reversal potential remains unchanged when bath solution is switched from 150 mM Na^+ to 150 mM K^+ . All data points are mean \pm s.e.m. ($n \geq 5$).

nomenclature as other voltage-gated channels, we labelled the six membrane-spanning helices within each 6-TM domain as IS1–IS6 and IIS1–IIS6, respectively (Extended Data Fig. 1). The overall structure of each 6-TM domain resembles that of the prokaryotic Na_v channels^{26,27} and contains two pore helices (P1 and P2) between S5 and S6 (Extended Data Figs 1 and 3). The AtTPC1 pore displays pseudo fourfold symmetry and superimposes well with other tetrameric channel pores (Extended Data Fig. 3d, e). However, this symmetry breaks down at the peripheral S1–S4 voltage-sensing domains (VSDs), which are attached to the pore with different relative positions within each subunit (Fig. 2d), resulting in a rectangular shaped channel dimer when viewed from the luminal side, with the two intra-subunit VSDs being more proximal than the inter-subunit VSDs (Fig. 2c). Notably, the relative position of VSD1 attachment to the pore of AtTPC1 resembles that of NavRh²⁷ whereas VSD2 is similar to NavAb²⁶ (Extended Data Fig. 3b, c). The EF-hand domain contains two tandem EF-hand motifs

and is located below VSD1 (Fig. 2b). The E1 helix of the first EF-hand comes from the C-terminal part of an exceptionally long IS6 helix; this structural feature allows for the Ca^{2+} -dependent conformational change at the EF-hand domain to be directly coupled to the pair of pore-lining IS6 helices in a functional channel.

Ion-conduction pore of AtTPC1

The AtTPC1 ion-conduction pore contains two pore helices between the outer (S5) and inner (S6) helices similar to prokaryotic Na_v channels^{26,27} (Fig. 3a). The pore is probably in a closed state since the four pore-lining inner helices form a bundle crossing at the cytosolic side with multiple constriction points that prevent the passage of hydrated cations (Fig. 3c and Extended Data Fig. 4a, b).

Unlike a K^+ channel filter, which forms a long narrow ion passageway with four well-defined ion binding sites for dehydrated K^+ , AtTPC1 has a much shorter and wider selectivity filter comprising residues ₂₆₄T₂₆₅

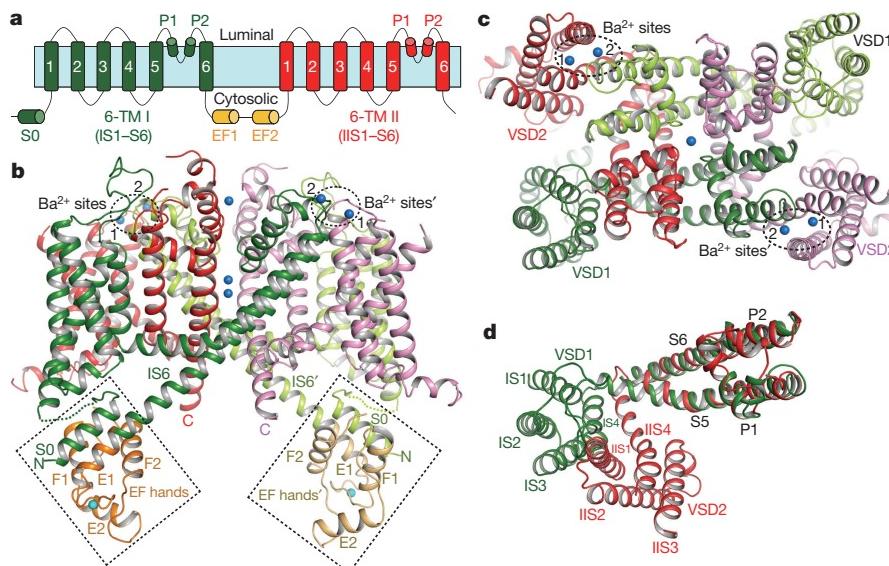


Figure 2 | Overall structure of AtTPC1. **a**, Topology diagram of AtTPC1. **b**, Side view of an AtTPC1 channel dimer. 6-TM I, 6-TM II, and EF hands from one subunit are shown in green, red and orange, respectively, and from the other symmetry related subunit are shown in lime green, purple, and light orange, respectively. The cytosolic EF-hand domains with bound

Ca^{2+} (cyan sphere) in EF1 are boxed and the two luminal Ba^{2+} (blue spheres) labelled 1 and 2 binding sites are circled. **c**, AtTPC1 viewed from luminal side. **d**, Superposition between the two 6-TM domains using the pore domains in the alignment. The orientation of 6-TM I is the same as in **c**.

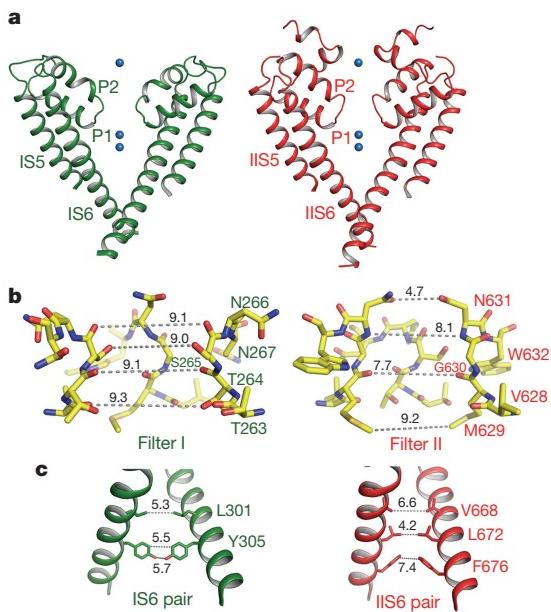


Figure 3 | The ion-conduction pore. **a**, The ion-conduction pore comprised of IS5–6 (left, green) and IIS5–6 (right, red). Ba²⁺ ions are shown as blue spheres. **b**, Structures of the selectivity filter formed by filter I (left) and filter II (right). **c**, Side view of the bundle crossing formed by IS6 pair (left) and IIS6 pair (right). Numbers are diagonal distances (in Å) of the constriction points.

from filter I and $\text{G}_{629}\text{MGN}_{631}$ from filter II (Fig. 3b). These filter residues surround the ion conduction pathway with both side-chain hydroxyl groups and main-chain carbonyls. The overall main-chain conformations of both filters, especially filter II, are similar to that of prokaryotic Na_v channels²⁶ (Extended Data Fig. 4c–e). The atom-to-atom cross distances along the major part of the filter ion pathway are around 8–9 Å (Fig. 3b). The side chain of Asn631 forms the narrowest point at the external entrance of the filter II cross-section with a distance of ~5 Å. However, Asn631 does not interact with any nearby residues and its side chain can freely rotate away from the central axis, rendering it unlikely to constrict ion permeation. The wide filter dimension in AtTPC1 implies that permeable ions cross the filter in a hydrated or partially hydrated state.

The crystallization condition for AtTPC1 also contained high concentrations of BaCl₂, and multiple Ba²⁺ ions were identified in the structure—three of which bind along the central pore axis: one at the external vestibule and two in the central cavity (Fig. 3a and Extended Data Fig. 4f). Unlike K⁺ channels, no Ba²⁺ is observed within the filter. Owing to the resolution limit, no clear electron density from ions or water molecules could be defined within the filter despite the presence of Na⁺, Ba²⁺ and Ca²⁺ in the crystallization conditions. Thus, a higher-resolution structure is required to define how permeable ions interact with the filter residues.

Cytosolic Ca²⁺ activation site

The AtTPC1 EF-hand domain follows the IS6 inner helix and contains two tandem EF-hand motifs (EF1 and EF2) where cytosolic Ca²⁺ binds and potentiates voltage activation (Figs 2a, b and 4a). Despite the presence of high Ba²⁺ concentrations in the crystallization conditions, no Ba²⁺ binding was observed in either EF hand, indicating high Ca²⁺ specificity. With the presence of 1 mM Ca²⁺ in the crystallization conditions, EF1 adopts a canonical Ca²⁺-bound EF-hand structure. The bound Ca²⁺ was also confirmed by anomalous scattering calculated from X-ray diffraction data collected at 2 Å wavelength using a crystal grown in the absence of Ba²⁺ (Fig. 4c). EF2, however, adopts an apo state, probably owing to a lower Ca²⁺ affinity, and its structure differs significantly from the canonical Ca²⁺-bound EF-hand. The E2 helix is distal from the F2 helix and the Ca²⁺-binding loop adopts

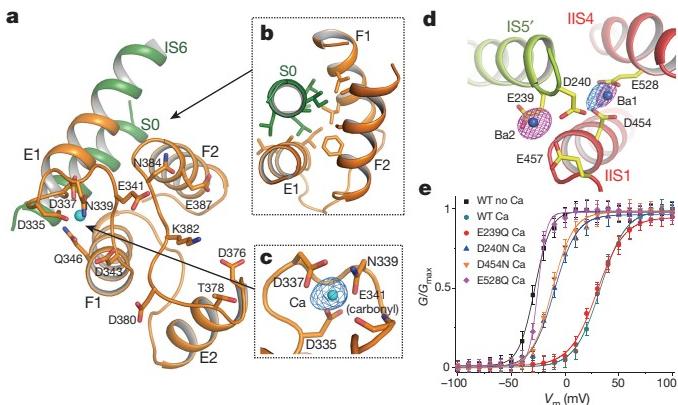


Figure 4 | The calcium modulation sites. **a**, Overall structure of the EF-hand domain with S0 and the C-terminal part of IS6 in green, and EF-hand helices in orange. Side chains are from residues predicted to participate in Ca²⁺ binding in EF1 and EF2. **b**, Packing interactions between S0 and E1/F1/F2. Residues contributing to the extensive hydrophobic contacts are: A34, L37, V38, L40, A41 and I45 on S0; A330, L333 and I334 on E1; L350 and L354 on F1; F388, C392 and A396 on F2. **c**, EF1 Ca²⁺ (cyan sphere) coordination with anomalous difference Fourier map (blue mesh contoured at 3.5σ). **d**, Luminal Ba²⁺ sites. Density from Ba²⁺ (magenta mesh at 11σ) and Ca²⁺ (blue mesh at 6σ) are defined by anomalous difference Fourier maps from native crystals grown with and without Ba²⁺, respectively. **e**, G/V curves of wild-type (WT) AtTPC1 and mutations at luminal Ba²⁺ sites recorded in the presence and absence of 100 μM extracellular Ca²⁺. Wild-type and mutant G/V curves recorded in the absence of Ca²⁺ are similar and only the wild-type one is shown. Data points are mean ± s.e.m. ($n \geq 5$).

an extended conformation. Consequently, those key Ca²⁺-binding residues are no longer properly positioned for Ca²⁺ coordination (Fig. 4a). Notably, a previous study on AtTPC1 demonstrated that only EF2 has an essential role in Ca²⁺ sensing²¹. This is also confirmed in our functional assay showing that a D335A mutation in the EF1 Ca²⁺ site retains cytosolic Ca²⁺ activation, whereas a D376A mutation in EF2 abolishes it (Extended Data Fig. 6a). Thus, only Ca²⁺ binding to EF2 triggers major conformational changes for channel activation and the structure of the EF-hand domain represents a deactivated state, despite the presence of Ca²⁺ at EF1. The tight protein packing around EF1 with the involvement of the S0 helix may explain the lack of EF1 Ca²⁺ activation. The N-terminal S0 helix of AtTPC1, although distal in primary sequence, is an integral part of the EF-hand domain and has been shown to be functionally indispensable²⁸. The S0 helix runs antiparallel to the E1 helix and is embedded in the deep hydrophobic groove formed by the E1, F1 and F2 helices (Fig. 4b). The extensive van der Waals interactions between S0 and EF1 probably lock the E1/F1 helices into a fixed position and prevent it from undergoing any structural change in response to Ca²⁺.

Luminal Ca²⁺ inhibition site

In contrast to cytosolic Ca²⁺, luminal Ca²⁺ is known to inhibit channel activation and Asp454 was previously identified to be important for luminal Ca²⁺ binding from a gain-of-function mutant *fou2* (refs 22–24). Two Ba²⁺ ions are observed in the vicinity of Asp454 (Fig. 4d). The site 1 Ba²⁺ is coordinated by the side-chain carboxylates of Asp454 on IIS1, Glu528 on IIS4, and Asp240 on IS5 from a neighbouring subunit. The second Ba²⁺ site is surrounded by residues Glu239, Asp240 and Glu457. Since Ba²⁺ exerts a similar inhibitory effect as Ca²⁺, albeit with weaker affinity (Extended Data Fig. 2b), they probably share the same inhibitory site. Two observations suggest that site 1 is the bona fide Ca²⁺ inhibition site and that the second Ba²⁺ binding is probably a consequence of high Ba²⁺ concentrations in the crystallization conditions. First, the anomalous difference map of a crystal grown in the absence of Ba²⁺ revealed a Ca²⁺ anomalous scattering peak at site 1 but not at site 2 (Fig. 4d). Second, neutralization mutations of the three

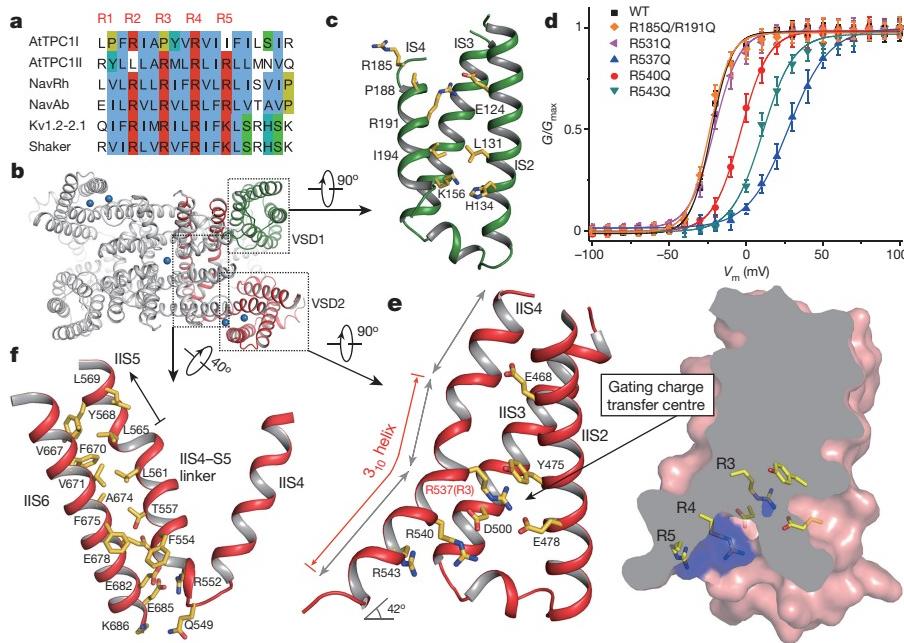


Figure 5 | The voltage-sensing domains. **a**, Partial S4 sequence alignment and arginine registry. **b**, Locations of VSD1, VSD2 and IIS4–S5 linker regions in AtTPC1. **c**, Side view of VSD1 with S1 omitted for clarity (same for VSD2). **d**, G/V curves of wild-type (WT) AtTPC1 and neutralization mutations of arginines on IS4 and IIS4. Data points are mean \pm s.e.m. ($n \geq 5$). **e**, Structure of VSD2 (left) and its surface-rendered cross-section

(right). Grey double arrows indicate the three segments of the curved IIS4 helix. Arginine in the gating charge transfer centre is labelled in red. **f**, Acute-angled connection between IIS4 and IIS4–S5 linker and the extensive interactions between the linker and IIS6. For clarity, the channel in **b** is rotated approximately 40° around the indicated axis.

site 1 acidic residues profoundly mitigated luminal Ca^{2+} inhibition, whereas mutagenesis at site 2, that is, Glu239Gln, has no effect (Fig. 4e and Extended Data Fig. 5). It is important to note that only VSD2 in AtTPC1 is voltage sensitive and its S4 helix (IIS4) is the primary mobile component during voltage activation as discussed later. Thus, luminal Ca^{2+} stabilizes VSD2 in the resting state by tethering IIS4 to the static IIS1 helix and the pore-forming IS5 of the neighbouring subunit, which in turn hinders IIS4 movement in response to voltage changes, analogous to extracellular Zn^{2+} inhibition observed in the voltage-gated proton channel Hv1 (ref. 29).

Voltage sensing domains in AtTPC1

The same gating charge numbering used for Kv1.2-2.1 (ref. 30) is adopted in sequence and structure comparison of various VSDs (Fig. 5a). VSD1 from AtTPC1 and its S4–S5 linker have a structural arrangement similar to that of the activated VSD of NavRh²⁷ (Fig. 5a–c and Extended Data Fig. 3). However, VSD1 in AtTPC1 lacks a few key features seen in canonical voltage-gated channels: the IS4 helix of AtTPC1 contains only two conserved arginine residues at R2 and R4; the 3_{10} -helix motif that is commonly seen in voltage-gated channels^{26,27,31,32} is not preserved in IS4, which forms a regular helix; and His and Leu respectively replace the highly conserved acidic and aromatic residues on S2 that form the charge transfer centre in voltage-gated channels³⁰, whereas Lys replaces the highly conserved acidic residue on S3. Consequently, VSD1 does not contribute to the voltage-dependent gating, and replacing both S4 arginines with neutral residues does not affect voltage activation of AtTPC1 (Fig. 5d and Extended Data Fig. 6b). VSD2 preserves the key elements of a canonical voltage sensor^{33–36} and is responsible for voltage-dependent gating in AtTPC1. The IIS4 helix contains four arginine residues, which corresponds to R1 (R531), R3 (R537), R4 (R540) and R5 (R543) (Fig. 5a) and mutagenesis analysis shows that R3 to R5, but not R1, contribute to voltage-sensing in AtTPC1 (Fig. 5d). Therefore, R537 at the R3 position represents the first gating charge in AtTPC1.

The majority of the IIS4 helix in VSD2 forms an exceptionally long, curved 3_{10} -helix from L533 to N547 (Fig. 5e), a feature initially

observed in the S4 helix of the MlotiK1 potassium channel³⁷ and, more recently, in several other voltage-gated channel structures^{26,27,31,32}. The bent IIS4 helix can be divided into three segments: the N-terminal segment preceding the 3_{10} -helix, the short middle segment from L533 to R537, and the long C-terminal segment after R537 running diagonally towards the intracellular membrane surface and connecting to the IIS4–S5 linker helix with a sharp turn (Fig. 5e, f and Extended Data Fig. 3a). The linker helix forms extensive interactions with IIS6, including salt bridges and hydrogen bonds at the beginning, followed by hydrophobic contact towards the end of the linker (Fig. 5f). This extensive interaction network ensures a coupled movement between the linker helix and IIS6 at the intracellular gate.

The structure of VSD2 is stabilized in a resting state by luminal Ba^{2+} and has several structural features distinct from other voltage-gated channels with an activated VSD. AtTPC1 IIS4 has its first gating charge (R537) positioned in the gating charge transfer centre³⁰, formed by highly conserved Y475/E478 from IIS2 and D500 from IIS3, whereas the activated VSD of NavAb²⁶ or Kv1.2-2.1 (ref. 31) has the last gating charge (R5 or K5) residing in the equivalent position (Fig. 5e and Extended Data Fig. 7). In AtTPC1, the long, curved C-terminal segment of IIS4 together with IIS1–S3 create a wide cavity below the charge transfer centre, allowing the rest of the voltage-sensing arginines (R4 and R5) to be exposed to the cytosol (Fig. 5e). However, in an activated VSD, the S4 segment below the charge transfer centre is a much shorter, straight helix, whereas the segment above is a longer, curved helix, which consequently narrows the cytosolic cavity but creates a deep, external aqueous cavity above the charge transfer centre where all the gating charges except the last one become exposed (Extended Data Fig. 7). This external cavity is occluded in the resting VSD2 of AtTPC1 since the N-terminal segment of IIS4 makes close contact with the external portion of IIS1–S3.

Voltage-gating mechanism

The structure of AtTPC1 provides a first glance of a voltage-gated channel in a resting state, allowing us to elucidate the structural basis of voltage sensing through structural comparison with NavAb²⁶.

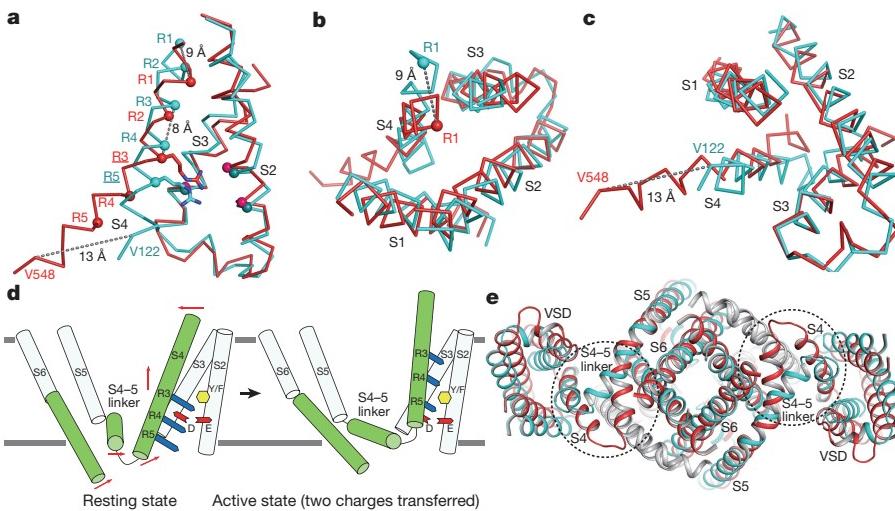


Figure 6 | Voltage gating mechanism. **a**, Side view of $C\alpha$ superposition between AtTPC1 VSD2 (red) and NavAb VSD (cyan) with S1 omitted. Spheres indicate the $C\alpha$ positions of critical residues for voltage sensing. Distances are between $C\alpha$ atoms of two equivalent S4 residues at the N terminus (R1–R1), middle (R3–R3) and C terminus (V122–V548). Arginines positioned in the gating charge transfer centre are underlined (R3 in AtTPC1 and R5 in NavAb). **b, c**, Luminal and cytosolic views of the

The S1–S3 regions of both channels align well, indicating that S1–S3 undergo no major movement during voltage sensing (Fig. 6a–c). A major difference between the two VSDs is the vertical positioning of their S4 helix. AtTPC1 has R3, whereas NavAb has R5, positioned in the charge transfer centre, representing a shift of approximately two β_{10} -helical turns. Attributable to the imperfect alignment between two different channels, the $C\alpha$ distance of about 8 Å between two equivalent gating charge residues (that is, R3s) is slightly less than two helical turns. In the context of AtTPC1 a sliding motion of two helical turns in S4 (~ 10 Å) from the resting (R3 in transfer centre) to the activated (R5 in transfer centre) state and resultant transfer of two gating charges across the membrane is plausible (Fig. 6d). The magnitude of the S4 movement and the total gating charges across the field probably vary among voltage-gated channels, depending on the number of gating charge residues. In NavAb with four and *Shaker* with five gating charges, voltage activation would give rise to three (~ 15 Å) and four helical-turn (~ 20 Å) displacements of S4, respectively, consistent with the estimation of 15–20 Å movement across the membrane in some studies^{38–40}. As most voltage-gated channels seen to date appear to have a β_{10} -helix at the gating charge region with all voltage-sensing arginines positioned in line with respect to one another, the screw-like helical rotation observed in the voltage-sensing phosphatase⁴¹ is unlikely to occur in the S4 helix of voltage-gated channels.

The S4 displacement during voltage-gating induces little conformational change in S3, indicating that S3b–S4 is unlikely to undergo a concerted paddle movement proposed from earlier studies on KvAP^{39,42}. This independent S4 movement is consistent with a recent study showing that removal of the complementarity between S3b and S4 in *Shaker* does not compromise voltage gating⁴³. Our study supports the conventional helix translation model^{44–46} but without rotation during voltage gating. However, the S4 helix does not move as a simple piston-like rigid unit. Its sliding movement is also accompanied by the bending of its N- and C-terminal segments, converting part of the vertical motion in the middle of S4 into lateral movement at the two S4 termini (Fig. 6a–d). Consequently, the N-terminal S4 segment seals off the external aqueous cavity in the resting state, while the C-terminal end of S4 undergoes more lateral movement on the internal membrane surface.

To visualize how S4 movement is coupled to the pore opening and closing, AtTPC1 is superimposed onto NavAb in the context of the

superposition, respectively. **d**, Cartoon representation of the translational S4 movement from the resting to activated states with two gating charges transferred. Red arrows indicate the directions of the movement at N-, middle, and C-terminal parts of S4, and at S4–5 linker and C-terminus of S6. **e**, Cytosolic view of the superposition between AtTPC1 (red) and NavAb (cyan) excluding the VSD1s of AtTPC1 and the equivalent VSDs of NavAb. Major structural changes highlighted in circles occur at S4 and S4–S5 linker.

whole channel with VSD1 omitted, as it does not contribute to voltage-gating and is positioned differently from the VSD of NavAb²⁶ (Fig. 6e). In AtTPC1, the downward IIS4 helix pushes the IIS4–S5 linker to tightly cuff around IIS6 at the bundle-crossing region, preventing the cytosolic gate from opening. In the activated NavAb, the upward S4 helix pulls the linker helix apart from S6. While the NavAb structure is defined as a pre-open state and its S6 helix appears to be decoupled from the linker, we expect the IIS6 inner helix to move concurrently with the IIS4–S5 linker upon voltage activation in AtTPC1 since its linker helix is tightly packed with IIS6 (Fig. 6d and Extended Data Fig. 8). In a tetrameric voltage-gated channel, this concerted movement of S6 and S4–S5 linker would dilate the gate. In AtTPC1, however, only VSD2 is voltage dependent, and the linker movement would only be coupled to the diagonal pair of IIS6 inner helices from the second 6-TM domains. We suspect that cytosolic Ca^{2+} binding in the EF-hand domains would introduce a similar kind of dilation movement to the other pair of IIS6 helices (Extended Data Fig. 8). This dual coupling mechanism explains the requirement for having both cytosolic Ca^{2+} and depolarization for AtTPC1 activation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 July; accepted 10 November 2015.

Published online 21 December 2015.

- Peiter, E. et al. The vacuolar Ca^{2+} -activated channel TPC1 regulates germination and stomatal movement. *Nature* **434**, 404–408 (2005).
- Calcraft, P. J. et al. NAADP mobilizes calcium from acidic organelles through two-pore channels. *Nature* **459**, 596–600 (2009).
- Hedrich, R. & Marten, I. TPC1-SV channels gain shape. *Mol. Plant* **4**, 428–441 (2011).
- Xu, H. & Ren, D. Lysosomal physiology. *Annu. Rev. Physiol.* **77**, 57–80 (2015).
- Rahman, T. et al. Two-pore channels provide insight into the evolution of voltage-gated Ca^{2+} and Na^+ channels. *Sci. Signal.* **7**, ra109 (2014).
- Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. Cloning of genomic and complementary DNA from *Shaker*, a putative potassium channel gene from *Drosophila*. *Science* **237**, 749–753 (1987).
- Ishibashi, K., Suzuki, M. & Imai, M. Molecular cloning of a novel form (two-repeat) protein related to voltage-gated sodium and calcium channels. *Biochem. Biophys. Res. Commun.* **270**, 370–376 (2000).
- Brailoiu, E. et al. Essential requirement for two-pore channel 1 in NAADP-mediated calcium signaling. *J. Cell Biol.* **186**, 201–209 (2009).
- Zong, X. et al. The two-pore channel TPCN2 mediates NAADP-dependent Ca^{2+} -release from lysosomal stores. *Eur. J. Phys.* **458**, 891–899 (2009).

10. Wang, X. et al. TPC proteins are phosphoinositide-activated sodium-selective ion channels in endosomes and lysosomes. *Cell* **151**, 372–383 (2012).
11. Cang, C. et al. mTOR regulates lysosomal ATP-sensitive two-pore Na^+ channels to adapt to metabolic state. *Cell* **152**, 778–790 (2013).
12. Cang, C., Bekele, B. & Ren, D. The voltage-gated sodium channel TPC1 confers endolysosomal excitability. *Nature Chem. Biol.* **10**, 463–469 (2014).
13. Jha, A., Ahuja, M., Patel, S., Brailoiu, E. & Mualem, S. Convergent regulation of the lysosomal two-pore channel-2 by Mg^{2+} , NAADP, PI(3,5)P₂ and multiple protein kinases. *EMBO J.* **33**, 501–511 (2014).
14. Pitt, S. J., Lam, A. K., Rietdorf, K., Galione, A. & Sitsapesan, R. Reconstituted human TPC1 is a proton-permeable ion channel and is activated by NAADP or Ca^{2+} . *Sci. Signal.* **7**, ra46 (2014).
15. Ruas, M. et al. Expression of Ca^{2+} -permeable two-pore channels rescues NAADP signalling in TPC-deficient cells. *EMBO J.* **34**, 1743–1758 (2015).
16. Sakurai, Y. et al. Ebola virus. Two-pore channels control Ebola virus host cell entry and are drug targets for disease treatment. *Science* **347**, 995–998 (2015).
17. Furuchi, T. et al. A putative two pore channel AtTPC1 mediates Ca^{2+} flux in *Arabidopsis* leaf cells. *Plant Cell Physiol.* **42**, 900–905 (2001).
18. Hedrich, R. & Neher, E. Cytoplasmic calcium regulates voltage-dependent ion channels. *Nature* **329**, 833–836 (1987).
19. Amodeo, G., Escobar, A. & Zeiger, E. A cationic channel in the guard cell tonoplast of *Allium cepa*. *Plant Physiol.* **105**, 999–1006 (1994).
20. Ward, J. M. & Schroeder, J. I. Calcium-activated K^+ channels and calcium-induced calcium release by slow vacuolar ion channels in guard cell vacuoles implicated in the control of stomatal closure. *Plant Cell* **6**, 669–683 (1994).
21. Schulze, C., Sticht, H., Meyerhoff, P. & Dietrich, P. Differential contribution of EF-hands to the Ca^{2+} -dependent activation in the plant two-pore channel TPC1. *Plant J.* **68**, 424–432 (2011).
22. Dadacz-Narloch, B. et al. A novel calcium binding site in the slow vacuolar cation channel TPC1 senses luminal calcium levels. *Plant Cell* **23**, 2696–2707 (2011).
23. Bonaventure, G. et al. A gain-of-function allele of TPC1 activates oxylipin biogenesis after leaf wounding in *Arabidopsis*. *Plant J.* **49**, 889–898 (2007).
24. Bonaventure, G., Gfeller, A., Rodriguez, V. M., Armand, F. & Farmer, E. E. The *fou2* gain-of-function allele and the wild-type allele of Two Pore Channel 1 contribute to different extents or by different mechanisms to defense gene expression in *Arabidopsis*. *Plant Cell Physiol.* **48**, 1775–1789 (2007).
25. Choi, W. G., Toyota, M., Kim, S. H., Hilleary, R. & Gilroy, S. Salt stress-induced Ca^{2+} waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proc. Natl Acad. Sci. USA* **111**, 6497–6502 (2014).
26. Payandeh, J., Scheuer, T., Zheng, N. & Catterall, W. A. The crystal structure of a voltage-gated sodium channel. *Nature* **475**, 353–358 (2011).
27. Zhang, X. et al. Crystal structure of an orthologue of the NaChBac voltage-gated sodium channel. *Nature* **486**, 130–134 (2012).
28. Larisch, N., Schulze, C., Galione, A. & Dietrich, P. An N-terminal dileucine motif directs two-pore channels to the tonoplast of plant cells. *Traffic* **13**, 1012–1022 (2012).
29. Takeshita, K. et al. X-ray crystal structure of voltage-gated proton channel. *Nature Struct. Mol. Biol.* **21**, 352–357 (2014).
30. Tao, X., Lee, A., Limapichat, W., Dougherty, D. A. & MacKinnon, R. A gating charge transfer center in voltage sensors. *Science* **328**, 67–73 (2010).
31. Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K^+ channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
32. Vieira-Pires, R. S. & Morais-Cabral, J. H. β_{10} helices in channels and other membrane proteins. *J. Gen. Physiol.* **136**, 585–592 (2010).
33. Armstrong, C. M. & Bezanilla, F. Charge movement associated with the opening and closing of the activation gates of the Na channels. *J. Gen. Physiol.* **63**, 533–552 (1974).
34. Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the *Shaker K⁺* channel. *Neuron* **16**, 1159–1167 (1996).
35. Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the *Shaker K⁺* channel. *Neuron* **16**, 1169–1177 (1996).
36. Schoppa, N. E., McCormack, K., Tanouye, M. A. & Sigworth, F. J. The size of gating charge in wild-type and mutant *Shaker* potassium channels. *Science* **255**, 1712–1715 (1992).
37. Clayton, G. M., Altieri, S., Heginbotham, L., Unger, V. M. & Morais-Cabral, J. H. Structure of the transmembrane regions of a bacterial cyclic nucleotide-regulated channel. *Proc. Natl Acad. Sci. USA* **105**, 1511–1515 (2008).
38. Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the *Shaker K⁺* channel S4. *Neuron* **16**, 387–397 (1996).
39. Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. The principle of gating charge movement in a voltage-dependent K^+ channel. *Nature* **423**, 42–48 (2003).
40. Ruta, V., Chen, J. & MacKinnon, R. Calibrated measurement of gating-charge arginine displacement in the KvAP voltage-dependent K^+ channel. *Cell* **123**, 463–475 (2005).
41. Li, Q. et al. Structural mechanism of voltage-dependent gating in an isolated voltage-sensing domain. *Nature Struct. Mol. Biol.* **21**, 244–252 (2014).
42. Jiang, Y. et al. X-ray structure of a voltage-dependent K^+ channel. *Nature* **423**, 33–41 (2003).
43. Xu, Y., Ramu, Y. & Lu, Z. *Shaker K⁺* channel with a miniature engineered voltage sensor. *Cell* **142**, 580–589 (2010).
44. Bezanilla, F. Voltage sensor movements. *J. Gen. Physiol.* **120**, 465–473 (2002).
45. Gandhi, C. S. & Isacoff, E. Y. Molecular models of voltage sensing. *J. Gen. Physiol.* **120**, 455–463 (2002).
46. Horn, R. Coupled movements in voltage-gated ion channels. *J. Gen. Physiol.* **120**, 449–453 (2002).

Acknowledgements We thank N. Nguyen for manuscript preparation, R. Hedrich at University of Würzburg and M. X. Zhu at University of Texas Health Science Center at Houston for providing clones of plant and animal TPC genes. The experimental results reported in this article derive from work performed at Argonne National Laboratory, Structural Biology Center (19ID) and GM/CA (23ID) at the Advanced Photon Source, and from work performed at the Berkeley Center for Structural Biology at the Advanced Light Source (ALS). Argonne is operated by UChicago Argonne, LLC, for the US Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. This work was supported in part by the Howard Hughes Medical Institute and by grants from the National Institutes of Health (GM079179 to Y.J.; NS055293 and NS074257 to D.R.) and the Welch Foundation (Grant I-1578 to Y.J.). The authors declare no competing financial interests.

Author Contributions J.G. performed the structure determination; W.Z., C.C. and D.R. performed electrophysiology; Q.C., C.L., L.C. and Y.Y. participated in sample preparation; J.G., W.Z. and Y.J. designed the research, analysed data, and prepared the manuscript.

Author Information The atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession number 5E1J. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.J. (youxing.jiang@utsouthwestern.edu).

METHODS

No statistical methods were used to predetermine sample size.

Protein expression and purification. The full-length *AtTPC1* gene (NCBI accession: NM_116594) was ligated into the pPICZ vector (Invitrogen) containing a C-terminal eGFP-8 × histidine tag. The plasmid was linearized with PmeI restriction enzyme and transformed into *Pichia pastoris* SMD1163 strain by electroporation (Bio-Rad). The positive strains integrated with recombinant *AtTPC1* gene were selected on agar plates containing 500 µg ml⁻¹ zeocin (Invitrogen). For protein expression, the transformed cells were grown in MGYH medium to an OD₆₀₀ = 3.0 and then induced in MMH medium for 2 days at 28 °C. Cells were harvested and washed in buffer A (50 mM Tris pH 8.0, 150 mM NaCl, 1 mM CaCl₂), and then frozen and stored at -80 °C until use.

The cells were re-suspended in buffer A and homogenized with an M-110P homogenizer (Microfluidics) for four times at 25,000 p.s.i. Whole-cell lysate was centrifuged at 10,000g for 10 min and the supernatant was subjected to a second round of centrifugation at 40,000 r.p.m. (Beckman type 45 Ti rotor) for 1 h to pellet the membrane. The membrane fraction was re-suspended in buffer A and homogenized with a glass dounce homogenizer. AtTPC1 was extracted using *n*-dodecyl-β-D-maltopyranoside (DDM) (Anatrace) at a concentration of 1% (w/v), stirring at 4 °C for 3 h. The supernatant, after extraction, was collected after a 40-min centrifugation at 48,000g at 4 °C and loaded onto a Talon cobalt affinity column (Clontech) followed by a wash of the column with three column volumes of buffer A + 0.05% (w/v) DDM + 20 mM imidazole. The detergent was then exchanged from DDM to 0.05% (w/v) lauryl maltose neopentyl glycol (LMNG) (Anatrace) on column by gravity flow. AtTPC1 was released from the column after histidine-tag removal by on-column thrombin digestion (Roche Diagnostics) at 4 °C overnight. The protein eluate was concentrated (100 kilodaltons MWKO centrifuge concentrator, Millipore) to 5–10 mg ml⁻¹ and further purified by size-exclusion chromatography (Superdex 200 column, GE Healthcare) in buffer SEC (20 mM Tris pH 8.0, 150 mM NaCl, 1 mM CaCl₂ and 0.05% (w/v) LMNG). The major peak eluted at around 11.2 ml and was pooled and concentrated to 5–14 mg ml⁻¹ for crystallization.

To obtain phasing information and to facilitate model building, more than 20 AtTPC1 mutants with a single-cysteine substitution at various parts of the protein were generated using Quickchange Site-Directed Mutagenesis Kit (Agilent). All mutants were expressed, purified and crystallized in similar conditions as the wild-type protein. Eighteen mutant proteins yielded crystals that diffracted X-rays to ~4.0 Å resolution.

Crystallization and data collection. The crystals were grown at 20 °C using conventional sitting drop vapour diffusion methods. Crystals appeared within 1–3 days in a condition consisting of 26% PEG400, 150 mM BaCl₂, 100 mM HEPES pH 7.0 and grew to their full size (0.1 mm × 0.2 mm × 0.2 mm) within 1–2 weeks. To identify the calcium-specific binding sites, crystals were also obtained in solutions containing 22% PEG400, 150 mM NaCl, 1 mM CaCl₂, 100 mM HEPES pH 7.0, and were used for data collection at longer wavelength (2 Å). For cryo-protection, the PEG400 in the reservoir solution was increased to 38% and crystals were allowed to equilibrate for 1 day before freezing in liquid nitrogen. The mercury-derivatized crystals were obtained by soaking the crystals in 0.5–1.0 mM CH₃HgCl for about 12 h before freezing.

X-ray diffraction data was collected using synchrotron radiation source (Advanced Photon Source 23IDB, 23IDD and 19ID; Advanced Light Source BL8.2.1 and BL8.2.2). The crystal belongs to space group C222₁ with cell dimensions of $a = 88.4 \text{ \AA}$, $b = 158.9 \text{ \AA}$, $c = 217.2 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$, and contains one subunit per asymmetric unit. The molecular dyad of a functional channel dimer coincides with the crystallographic dyad. To maximize the anomalous signal, the mercury-derivative data were collected near the mercury absorption edge ($\lambda = 1.0070 \text{ \AA}$), and the data of native crystal grown in Ca²⁺ (without Ba²⁺) were collected at $\lambda = 2.0000 \text{ \AA}$.

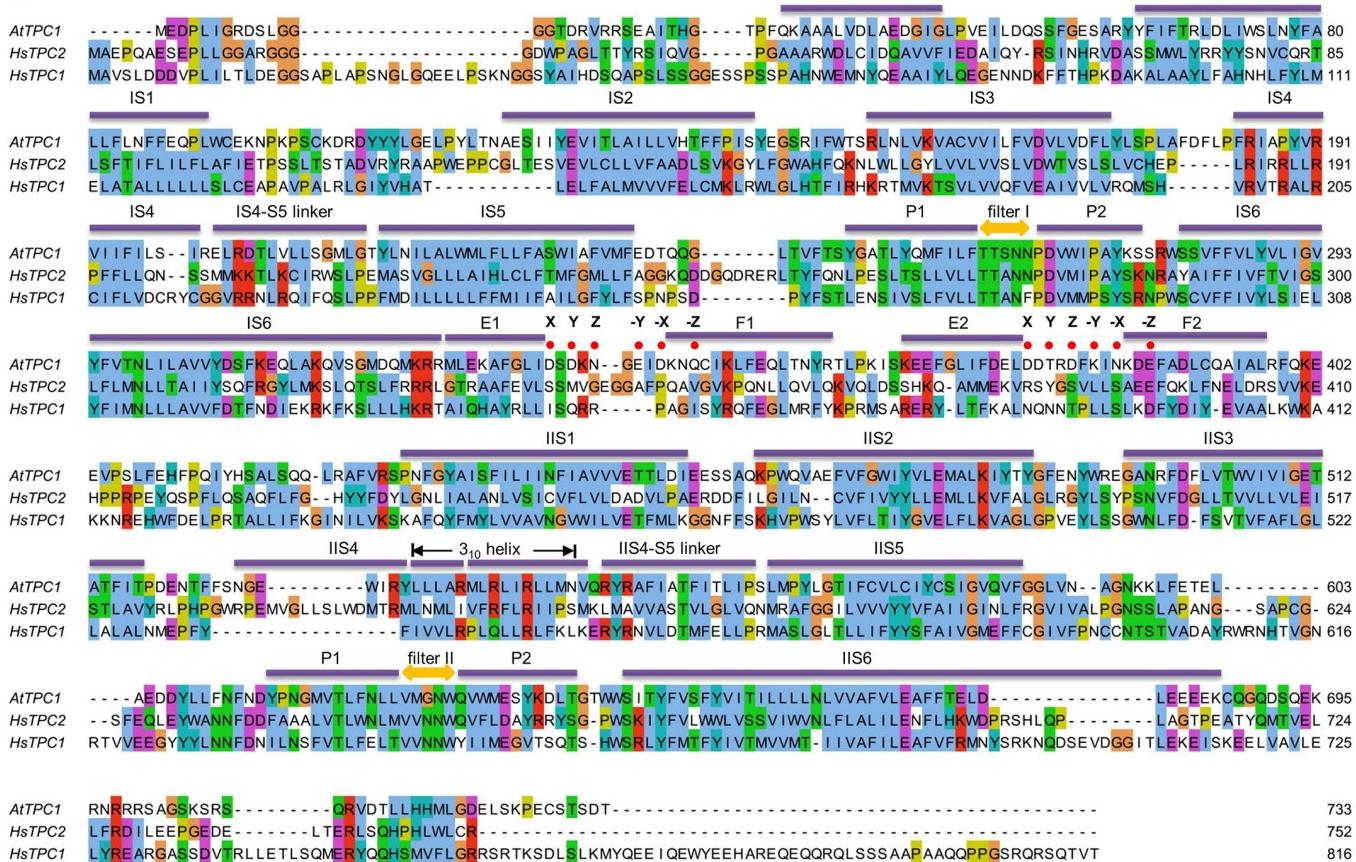
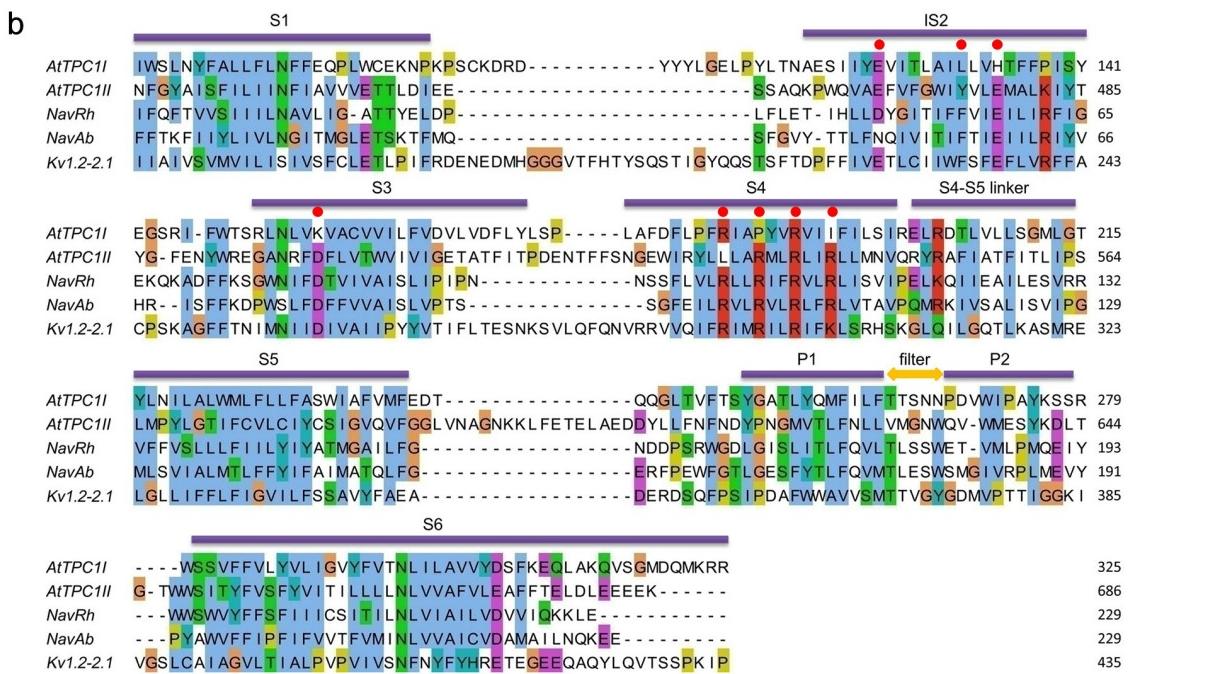
Structure determination. The diffraction data was integrated and scaled using the HKL2000 package⁴⁷. Since the diffraction data were anisotropic, ellipsoidal truncation, anisotropic scaling and B-factor sharpening were applied to the data using two approaches. In the first approach, the ‘auto-correction’ function in HKL2000 was applied during the final scaling. Data after auto-correction yielded better experimental maps (Extended Data Fig. 9a). However, a significant amount of data was discarded in this approach, resulting in a very low completeness at high-resolution shells. This approach was mainly used in the initial phasing and map calculation for model building. The anisotropy server (UCLA) was used in the second approach⁴⁸. In this approach, the best native data were truncated to 1/3.3 Å⁻¹, 1/4.1 Å⁻¹ and 1/3.5 Å⁻¹ along a^* , b^* and c^* , respectively. After anisotropic scaling, an isotropic B-factor of -63.9 Å² was applied. The data processed in the second approach have higher completeness and were used in the final refinement.

The structure was determined by single isomorphous replacement with anomalous scattering (SIRAS). The native data and the mercury-derivatized A604C mutant data were used to calculate the experimental phases using the AutoSharp Suite⁴⁹. The heavy-atom positions were determined in SHELXD⁵⁰ and refined in SHARP^{51,52}. The initial phases were improved by solvent flattening with SOLOMON⁵³. The experimental electron density map is of sufficient quality for initial assignment of most helical elements of the channel (Extended Data Fig. 9a). To facilitate accurate model building, we also obtained 14 mutant crystals containing one single-cysteine substitution at various parts of the protein. These mutant crystals were also derivatized by soaking with CH₃HgCl, which, together with the heavy-atom sites from the wild-type crystals, provided unambiguous registry for 20 residues throughout the protein, allowing us to accurately model the structured regions of AtTPC1 (Extended Data Fig. 9b). PHENIX⁵⁴ and Coot⁵⁵ were used for the refinement and model building, respectively. As there are several barium ions in the native structure, F(+) and F(−) were separated in the data used for refinement. The final structure was refined to 3.3 Å with R_{work} of 32.5% and R_{free} of 33.2%, and contained residues 32–53, 62–173, 184–402, 415–518, 524–590 and 595–686, covering 84% of the full-length AtTPC1. The geometry of the final structural model was analysed with Procheck⁵⁶, giving statistics of 90.2%, 9.6%, 0.2% and 0.0% for the most favoured, additional allowed, generously allowed and disallowed regions, respectively, on the Ramachandran plot. The bound Ca²⁺ ions at EF1 and the luminal inhibition site were confirmed by calcium anomalous scattering. The anomalous difference Fourier map was calculated from 4 Å resolution X-ray diffraction data collected at 2 Å wavelength using a crystal grown in the absence of Ba²⁺. The data collection and refinement statistics are listed in Extended Data Table 1. All the structure figures in this paper were prepared with PyMOL⁵⁷.

Electrophysiology. The *AtTPC1* open reading frame (ORF) was cloned into SalI/SmaI sites of the pEGFP-C1 vector (Clontech). All single-site mutants were generated using Quikchange Site-Directed Mutagenesis Kit (Agilent) and confirmed by DNA sequencing. 1–2 µg of the plasmid was transfected into HEK293 cells that were grown as a mono-layer in 35-mm tissue culture dishes (to ~70% confluence) using Lipofectamine 2000 (Life Technology). 24–48 h after transfection, cells were dissociated by trypsin treatment and kept in complete serum-containing medium and re-plated onto 35-mm tissue culture dishes and incubated in a tissue culture incubator until recording. Patch clamp in the whole-cell configuration was employed to measure AtTPC1 current in HEK293 cells expressing GFP-AtTPC1. The standard bath solution contained (in mM): 145 sodium methanesulfonate (Na-MS), 5 NaCl, 10 HEPES buffered with Tris, pH 7.4. The pipette solution contained (in mM): 150 Na-MS, 2.5 MgCl₂, 10 HEPES buffered with Tris, pH 7.4. For free Ca²⁺ concentrations less than 100 µM, a mixture of 5 mM EGTA and certain amount of CaCl₂ was prepared to achieve the target free Ca²⁺ concentration according to MAXCHELATOR (<http://maxchelator.stanford.edu>). The patch pipettes were pulled from borosilicate glass (Harvard Apparatus) and heat polished to a resistance of 3–5 MΩ. Data was acquired using an AxoPatch 200B amplifier (Molecular Devices) and a low-pass analogue filter set to 1 kHz. The current signal was sampled at a rate of 20 kHz using a Digidata 1322A digitizer (Molecular Devices) and further analysed with pClamp 9 software (Molecular Devices). After the patch pipette attached to the cell membrane, a giga seal (5–10 GΩ) was formed by gentle suction. The whole-cell configuration was formed by short electrical stimulation or suction to rupture the patch. The holding potential was set to -70 mV. The whole-cell current reached a maximum and remained stable within ~5 min. The membrane was stepped from the holding potential (-70 mV) to various testing potentials (-100 mV to +100 mV) for 1 s and then returned to the holding potential. The peak tail currents were used to generate G/G_{max} versus V curves ($G = I/V$). G_{max} in most experiments was obtained from the peak tail current at 100 mV testing potential with the presence of 300 µM [Ca²⁺]_{cytosolic} and 0 mM [Ca²⁺]_{extracellular}. $V_{1/2}$ and Z values were obtained from the fits of data with Boltzmann equation, where $V_{1/2}$ is the voltage at which the channels have reached half of their maximum fraction open and Z is the apparent valence of voltage dependence. To determine the selectivity of AtTPC1, the membrane potential was stepped to +80 mV for 1 s to fully activate the channels and then switched to various testing potentials (-100 mV to +60 mV). The tail currents were recorded to generate an I/V curve for the determination of the reversal potential. All data points are mean ± s.e.m. ($n \geq 5$).

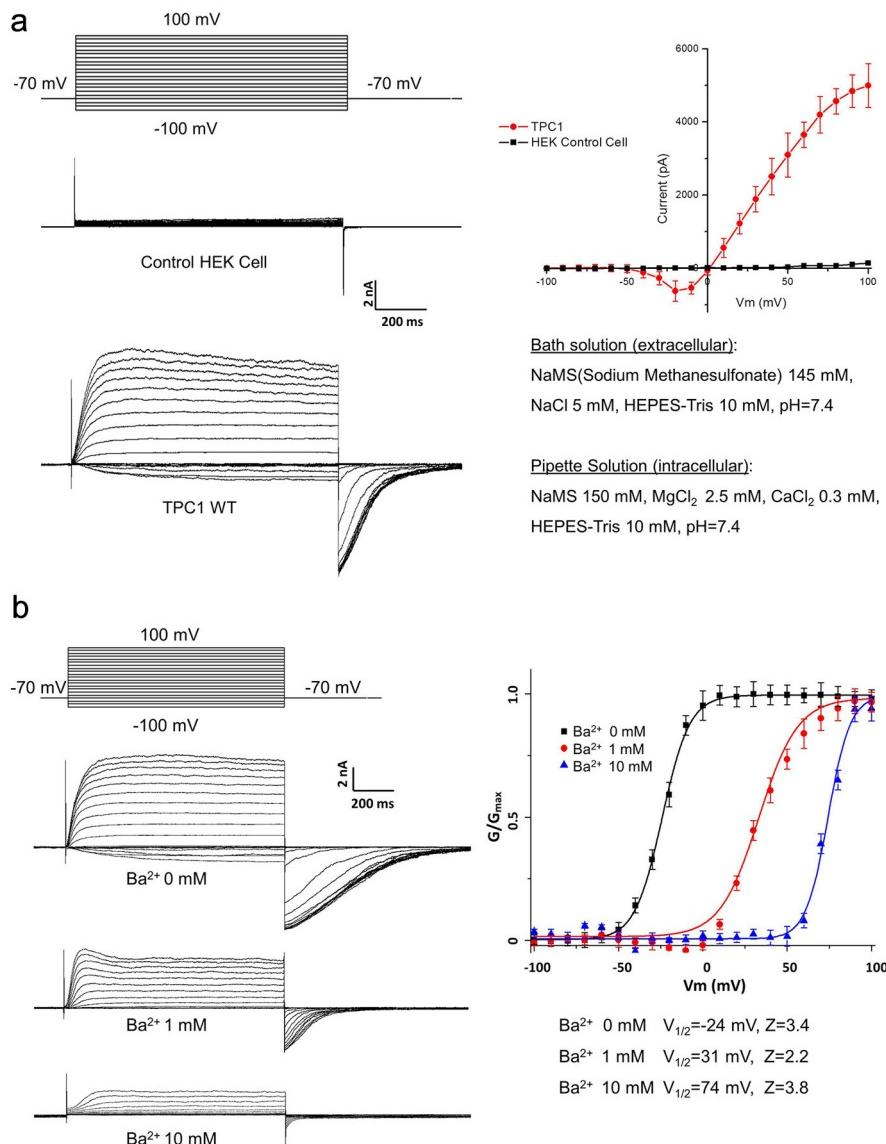
47. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
48. Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **103**, 8060–8065 (2006).
49. Vonrhein, C., Blanc, E., Roversi, P. & Brigitte, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).

50. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
51. Fortelle, E. I. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
52. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
53. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1-ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
54. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
56. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
57. The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

a**b**

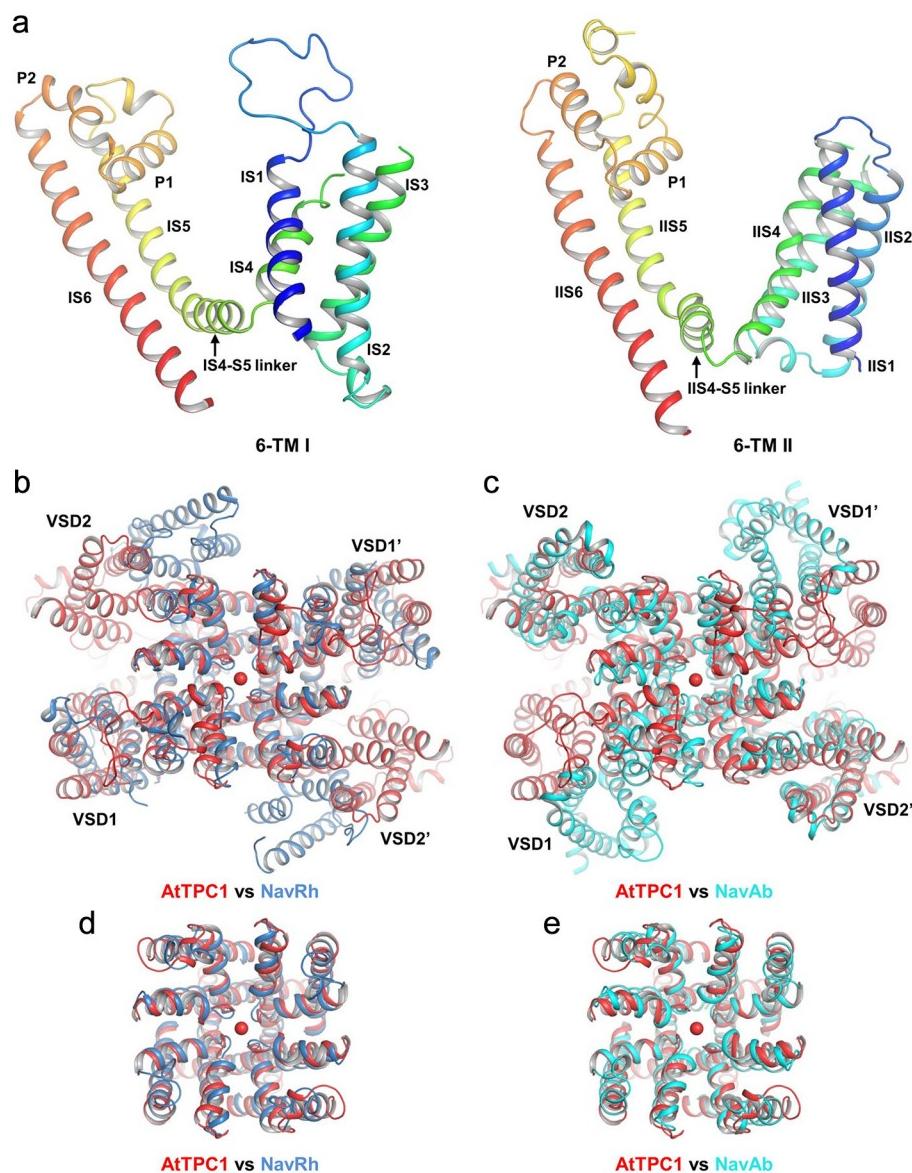
Extended Data Figure 1 | Sequence analysis. a, Sequence alignment of AtTPC1, human TPC1 (HsTPC1) and TPC2 (HsTPC2). Secondary structure assignments are based on the AtTPC1 structure. Red dots indicate the residues predicted to participate in calcium coordination in EF-hand domains. **b,** Sequence alignment of the two 6-TM domains of

AtTPC1 (AtTPC1I and AtTPC1II), NavRh (Protein Data Bank (PDB) accession: 4DXW), NavAb (PDB: 3RVY) and Kv1.2-2.1 (PDB: 2R9R). Red dots indicate the residues critical for voltage sensing. Secondary structure assignments are based on the AtTPC1 6-TM I structure.



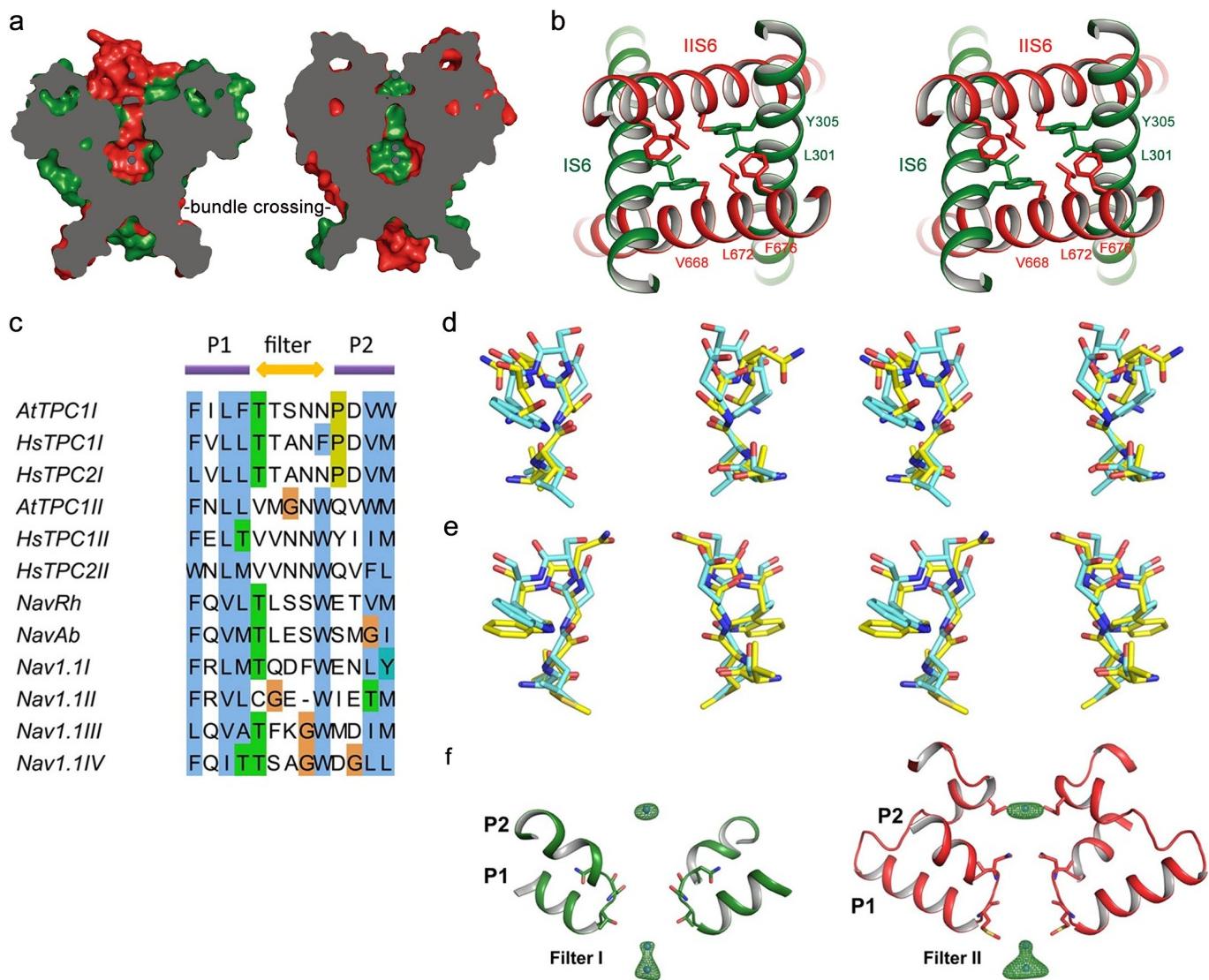
Extended Data Figure 2 | Voltage activation and Ba²⁺ modulation of AtTPC1 overexpressed in HEK cells. **a**, Voltage-dependent activation of wild-type AtTPC1. Channel currents were recorded using patch clamp in the whole-cell configuration. The membrane was stepped from holding potential (-70 mV) to various testing potentials and then returned to the

holding potential. The I/V curve was plotted using the steady peak current against the voltage. The peak tail currents were recorded to generate the G/V curves for voltage activation analysis. **b**, Extracellular Ba²⁺ inhibition of AtTPC1. The intracellular solution (pipette) contains 300 μ M Ca²⁺ necessary for channel activation.



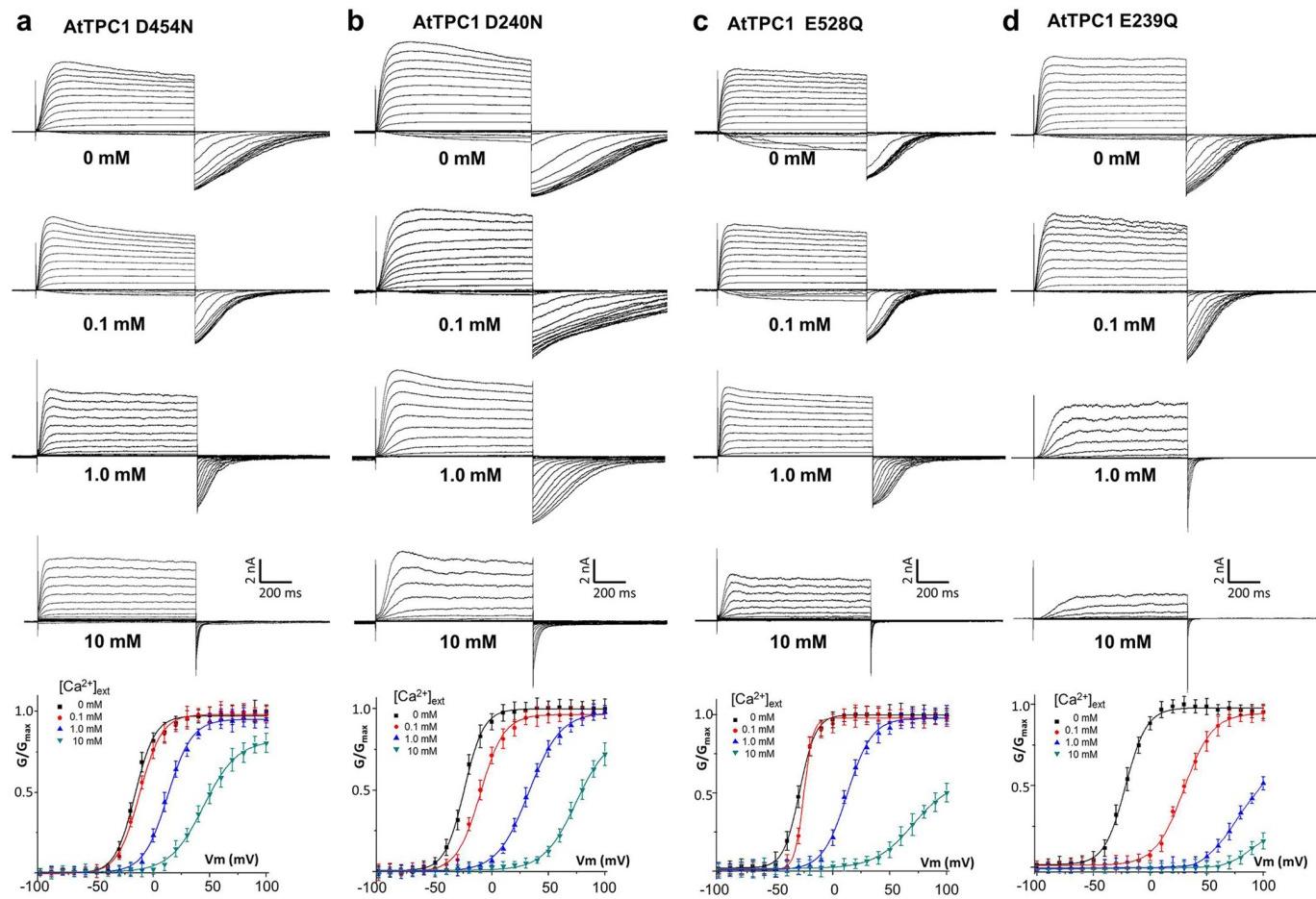
Extended Data Figure 3 | Structure of AtTPC1 transmembrane region and its alignment with prokaryotic Nav channels. **a**, Structure of the individual 6-TM domain of AtTPC1 in rainbow colour with the same pore orientation. **b**, Superposition of AtTPC1 (red) and NavRh (blue, PDB: 4DXW). The NavRh VSDs align well with AtTPC1 VSD1s.

c, Superposition of AtTPC1 (red) and NavAb (cyan, PDB: 3RVY). The NavAb VSDs align well with AtTPC1 VSD2s. **d**, Pore superposition between AtTPC1 (red) and NavRh (blue). **e**, Pore superposition between AtTPC1 (red) and NavAb (cyan).


Extended Data Figure 4 | The ion-conduction pore of AtTPC1.

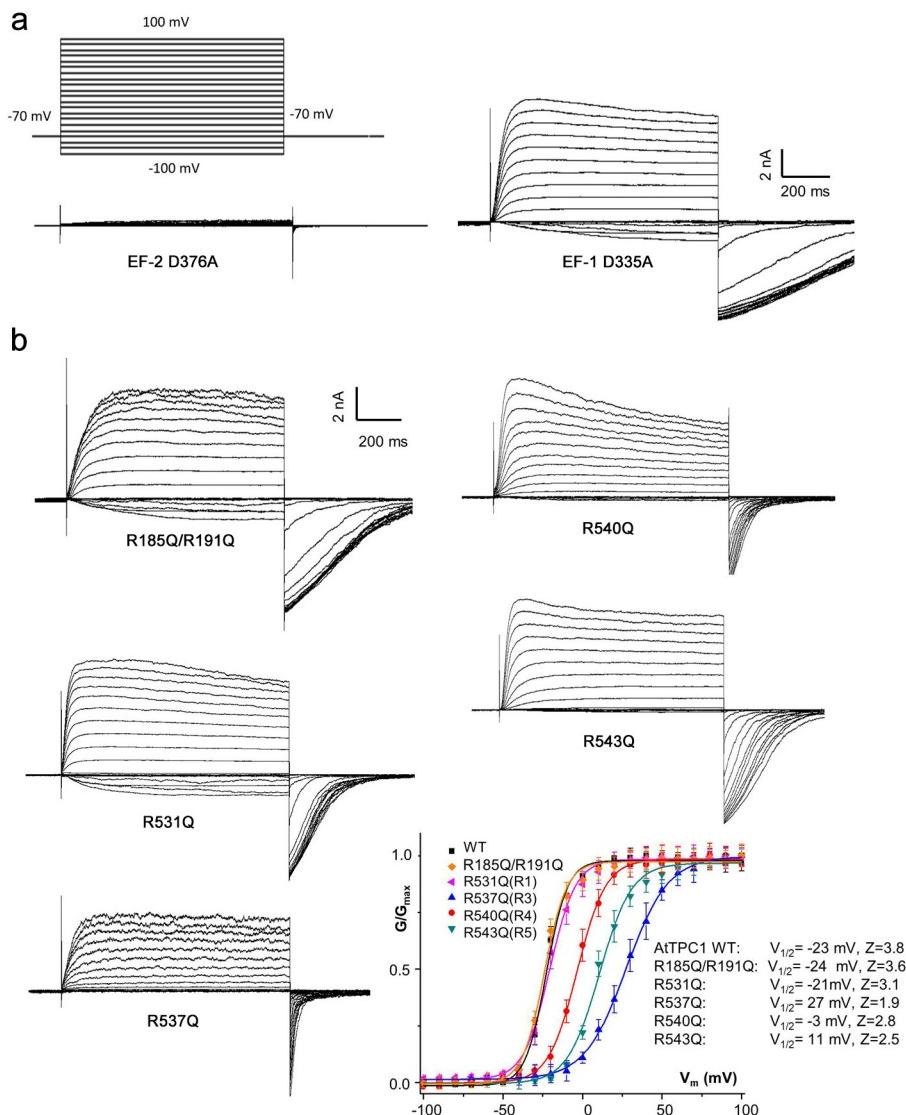
a, Cross-sections of surface-rendered AtTPC1 pore along IS6 pair (left) and IIS6 pair (right). The channel is closed at the bundle crossing. **b**, Stereo view of the bundle crossing region from the cytosolic side. **c**, Partial sequence alignment of the selectivity filters from two pore channels (AtTPC1, HsTPC1 and HsTPC2), bacterial sodium channels (NavRh and NavAb) and human voltage-gated sodium channel Nav1.1. **d**, Stereo view

of the structural alignment between AtTPC1 Filter I (carbon in yellow) and NavAb filter (carbon in cyan). **e**, Stereo view of structural alignment between AtTPC1 filter II and NavAb filter. **f**, Anomalous difference Fourier map of native crystal (green mesh, 4.5σ level) reveals the bound Ba^{2+} along the ion-conduction pathway. The two cavity sites are probably occupied by a single Ba^{2+} ion alternatively, as the two sites are only 3\AA apart, too close to accommodate two ions simultaneously.



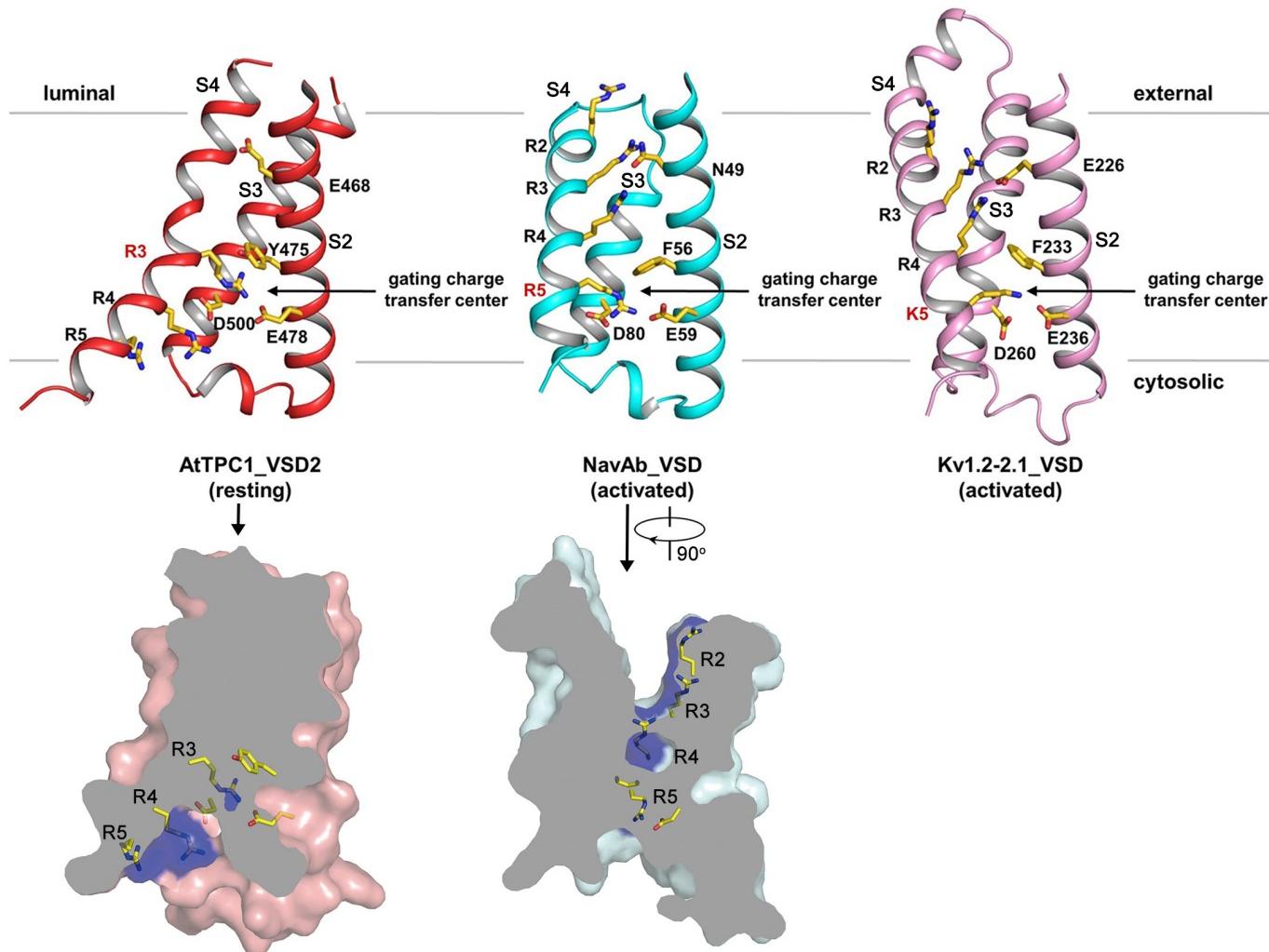
Extended Data Figure 5 | The whole cell currents and G/V curves of AtTPC1 with mutations at the luminal Ba^{2+} binding sites. a–d, D454N (a), D240N (b), E528Q (c) and E239Q (d). The bath solutions contained 0,

0.1, 1, or 10 mM $[\text{Ca}^{2+}]_{\text{extracellular}}$. The pipette solutions contained 300 μM $[\text{Ca}^{2+}]_{\text{cytosolic}}$. Data measured in 0.1 mM $[\text{Ca}^{2+}]_{\text{extracellular}}$ are shown in the main text Fig. 4e.



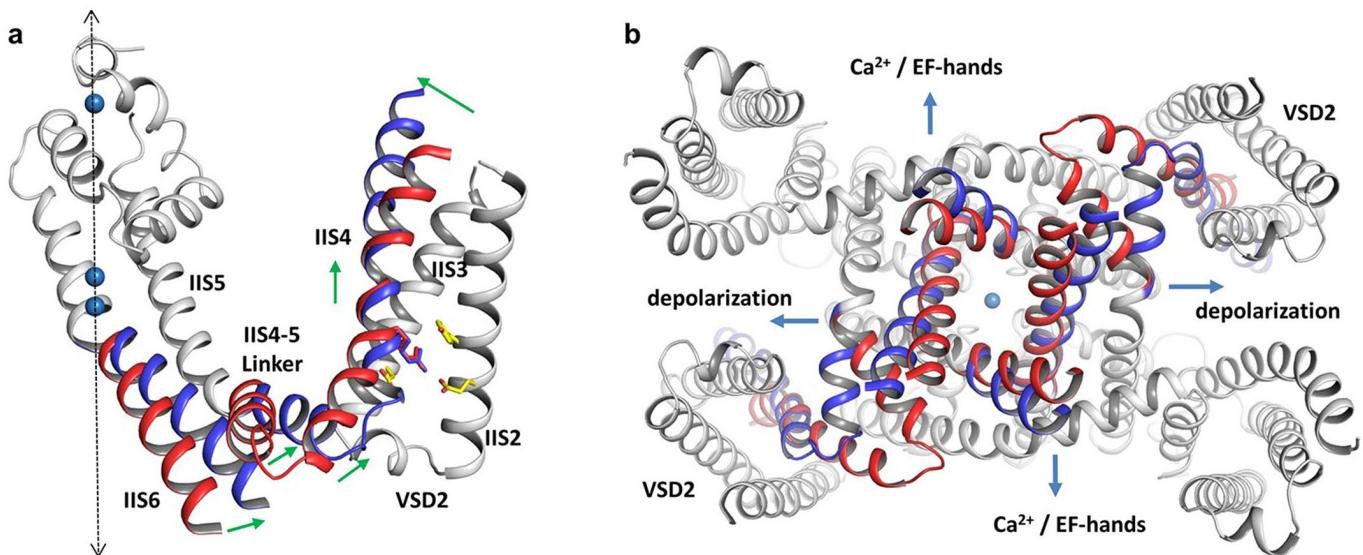
Extended Data Figure 6 | Functional analysis of AtTPC1 mutants.
a, The whole-cell currents of AtTPC1 containing EF-hand Ca^{2+} -site mutations (D335A in EF1 and D376A in EF2). Currents were recorded

with the presence of $300 \mu\text{M} [\text{Ca}^{2+}]_{\text{cytosolic}}$. **b**, Whole-cell currents and G/V curves of AtTPC1 with neutralization mutations of arginines on IS4 and IIS4 of the voltage-sensing domains.



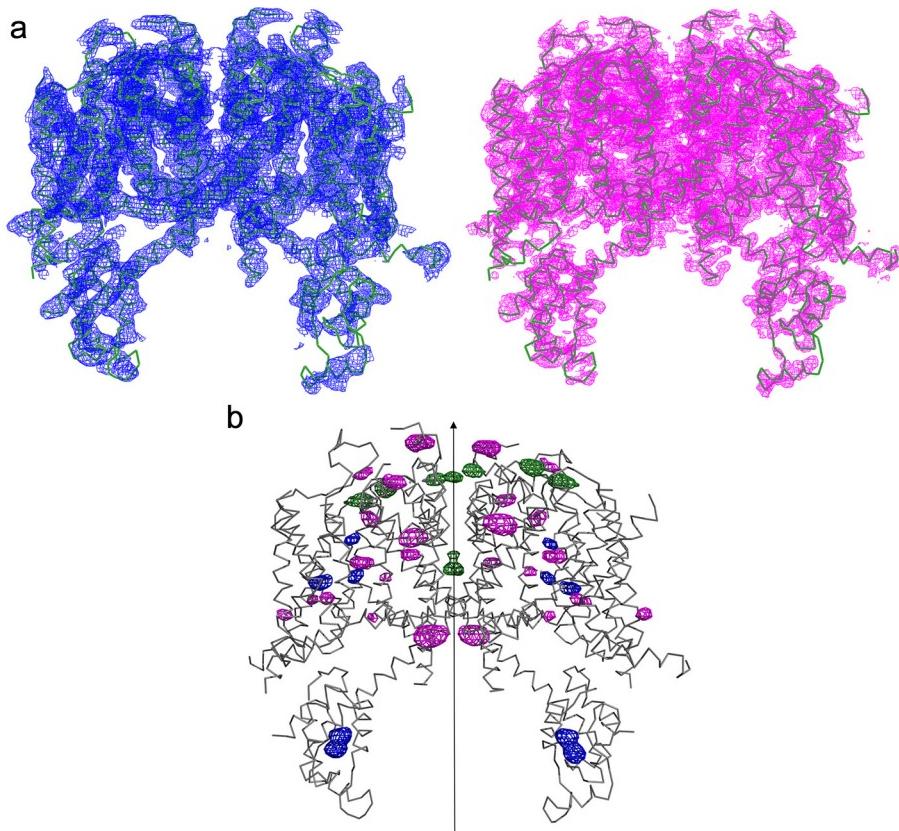
Extended Data Figure 7 | Structural comparison between AtTPC1 VSD2, NavAb VSD (PDB: 3RVY) and Kv1.2-2.1 VSD (PDB: 2R9R). All structures are aligned at the gating charge transfer centre and S1 helices are removed for clarity. The side chains of the voltage-sensing arginines in S4, residues in gating charge transfer centre and the conserved acidic

residue in S2 are shown in stick model. Voltage-sensing residues in gating charge transfer centre are labelled in red. Lower panels are cross-sections of surface-rendered AtTPC1 VSD2 (left) and NavAb VSD (right) with S4 gating charge arginines in blue. NavAb VSD is rotated by 90° to visualize the external aqueous cavity.



Extended Data Figure 8 | Proposed model for AtTPC1 activation. **a**, The model of AtTPC1 6-TM II in voltage-activated state is generated based on the structural comparison between AtTPC1 and NavAb. Only IIS4, IIS4–S5 linker and IIS6 are considered as the moving parts, assuming IIS6 moves concurrently with IIS4–S5 linker. The moving parts are coloured red for resting state and blue for activated state. The rest of the protein is coloured in grey. Green arrows indicate the directions of the movement at

the N terminus, middle part, and C terminus of IIS4, and at IIS4–S5 linker and C terminus of IIS6. Dashed arrow indicates the central axis of the channel. **b**, Cytosolic view of the channel-opening mechanism. Compared with the closed state (red), membrane depolarization and calcium binding to EF-hand domain lead to the opening of IIS6 and IS6 (modelled in blue), respectively.



Extended Data Figure 9 | Structure determination of AtTPC1.

a, Experimental electron density maps superposed with the final refined model. Density in blue (left) is the experimental SIRAS map calculated from the native and mercury-derivative data without anisotropic truncation and B-factor sharpening. Density in magenta (right) is the experimental SIRAS map calculated from the same native and mercury-derivative data after anisotropic truncation and B-factor sharpening using 'auto correction' in HKL2000. This map provides much better structural features, that is, side chains. All maps are contoured at 1.5σ level.

b, Anomalous difference Fourier maps of mercury-derivatized native and

mutant crystals superposed on the final refined model. The blue density peaks indicate the positions of mercury bound to the native cysteine residues. The magenta density peaks indicate the positions of mercury bound to cysteine residues introduced into various part of the protein (single-cysteine mutants). The green density peaks are calculated from the wild-type crystal (no mercury soaking), indicating the barium positions in wild-type AtTPC1. All maps are contoured at 4σ . Total 20 residues in each subunit are accurately registered by the mercury sites. Arrow indicates the molecular dyad of the channel dimer.

Extended Data Table 1 | Data collection and refinement statistics

Dataset	Native	A604C_Hg	Ca_2A ^a
Data collection			
Space group	C222 ₁	C222 ₁	C222 ₁
Cell dimensions			
<i>a, b, c</i> (Å)	88.44, 158.85, 217.24	88.57, 158.19, 217.03	88.10, 151.00, 214.91
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.0332	1.0070	2.0000
Resolution (Å)	50.00-3.30 (3.36-3.30) ^b	50.00-3.30 (3.36-3.30)	50.00-4.00 (4.07-4.00)
R_{merge}	0.060 (0.809)	0.052 (>1.000)	0.051 (0.302)
$CC_{1/2}$	(0.924)	(0.878)	(0.927)
I/σ	36.1 (1.6)	24.4 (0.9)	25.4 (2.0)
Completeness (%)	96.2 (78.2)	94.7 (74.4)	81.8 (58.7)
Redundancy	6.5 (5.2)	5.5 (4.5)	6.4 (3.8)
Refinement			
Resolution (Å)	3.3 * 4.1 * 3.5 ^c		
No. reflections	34119		
$R_{\text{work}}/R_{\text{free}}$ ^d	0.3247/0.3321		
No. atoms			
Protein	4949		
Ligand/ion	11		
Water	4		
B-factors			
Protein	91.12		
Ligand/ion	120.39		
Water	55.33		
R.m.s deviations			
Bond lengths (Å)	0.006		
Bond angles (°)	0.854		

^aThe crystal was grown in 1 mM CaCl₂ and no barium; the data was collected at 2 Å wavelength to maximize the calcium anomalous signal.^bThe numbers in the parentheses show the values in the highest resolution shell.^cThe data was elliptically truncated to 3.3 × 4.1 × 3.5 Å along *a**, *b**, and *c**.^d R_{free} was calculated with 5% of reflection data.

A repeating fast radio burst

L. G. Spitler¹, P. Scholz², J. W. T. Hessels^{3,4}, S. Bogdanov⁵, A. Brazier^{6,7}, F. Camilo^{5,8}, S. Chatterjee⁶, J. M. Cordes⁶, F. Crawford⁹, J. Deneva¹⁰, R. D. Ferdman², P. C. C. Freire¹, V. M. Kaspi², P. Lazarus¹, R. Lynch^{11,12}, E. C. Madsen², M. A. McLaughlin¹², C. Patel², S. M. Ransom¹³, A. Seymour¹⁴, I. H. Stairs^{2,15}, B. W. Stappers¹⁶, J. van Leeuwen^{3,4} & W. W. Zhu¹

Fast radio bursts are millisecond-duration astronomical radio pulses of unknown physical origin that appear to come from extragalactic distances^{1–8}. Previous follow-up observations have failed to find additional bursts at the same dispersion measure (that is, the integrated column density of free electrons between source and telescope) and sky position as the original detections⁹. The apparent non-repeating nature of these bursts has led to the suggestion that they originate in cataclysmic events¹⁰. Here we report observations of ten additional bursts from the direction of the fast radio burst FRB 121102. These bursts have dispersion measures and sky positions consistent with the original burst⁴. This unambiguously identifies FRB 121102 as repeating and demonstrates that its source survives the energetic events that cause the bursts. Additionally, the bursts from FRB 121102 show a wide range of spectral shapes that appear to be predominantly intrinsic to the source and which vary on timescales of minutes or less. Although there may be multiple physical origins for the population of fast radio bursts, these repeat bursts with high dispersion measure and variable spectra specifically seen from the direction of FRB 121102 support an origin in a young, highly magnetized, extragalactic neutron star.^{11,12}

FRB 121102 was discovered⁴ in the PALFA survey, a deep search of the Galactic plane at 1.4 GHz for radio pulsars and fast radio bursts (FRBs) using the 305-m William E. Gordon Telescope at the Arecibo Observatory and the 7-beam Arecibo L-band Feed Array (ALFA)^{13,14}. The observed dispersion measure (DM) of the burst is roughly three times the maximum value expected along this line of sight in the NE2001 model¹⁵ of Galactic electron density, that is, $\beta_{\text{DM}} \equiv \text{DM}_{\text{FRB}} / \text{DM}_{\text{max}}^{\text{Gal}} \approx 3$, suggesting an extragalactic origin.

Initial Arecibo follow-up observations were limited in both dwell time and sky coverage and resulted in no detection of additional bursts⁴. In 2015 May and June we carried out more extensive follow-up using the Arecibo telescope, covering an $\sim 9'$ radius with a grid of six ALFA pointings around the then-best sky position of FRB 121102 (Fig. 1 and Extended Data Tables 1 and 2). As described in Methods, high-time-resolution, total intensity spectra were recorded, and the data were processed using standard radio-frequency interference (RFI) excision, dispersion removal, and single-pulse-search algorithms implemented in the PRESTO¹⁶ software suite and associated data reduction pipelines^{14,17}.

We detected ten additional bursts from FRB 121102 in these observations. The burst properties, and those of the initial FRB 121102 burst, are listed in Table 1. The burst intensities are shown in Fig. 2. No other periodic or single-pulse signals of a plausible astrophysical origin were detected at any other DM. Until the source's physical nature is clear, we continue to refer to it as FRB 121102

and label each burst chronologically starting with the original detection.

The ten newly detected bursts were observed exclusively in two adjacent sky positions of the telescope pointing grid located $\sim 1.3'$ apart (Fig. 1 and Extended Data Table 1). The unweighted average J2000 position from the centres of these two beams is right ascension $\alpha = 05^{\text{h}} 31^{\text{m}} 58^{\text{s}}$ and declination $\delta = +33^{\circ} 08' 04''$, with an uncertainty radius of about $3'$. The corresponding Galactic longitude and latitude are $l = 174.89^{\circ}$, $b = -0.23^{\circ}$. This more accurate position is $3.7'$ from the beam centre of the discovery burst⁴, meaning that FRB 121102 burst 1 was detected well off-axis, as originally concluded⁴.

The measured DMs of all 11 bursts are consistent to within the uncertainties, and the dispersion indices (dispersive delay $\Delta t \propto \nu^{-\xi}$, where ν is the radio frequency) match the $\xi = 2.0$ value expected for radio waves travelling through a cold, ionized medium. This is strong evidence that a single astronomical source is responsible for the events. In addition, the ~ 0.002 DM index uncertainty we calculate for burst 11 (see Methods) is slightly less than that reported for FRB 110523 (ref. 8), making this the most precise determination of dispersion index for any FRB thus far. The upper bound on the dispersion index is identical to that of FRB 110523 (ref. 8) and, hence, following the same arguments used there, burst 11 provides a similar lower limit of ten astronomical units (1 AU is the Earth–Sun distance) for the size of the dispersive region.

The 11 bursts have peak flux densities $gS_{1400} \approx 0.02\text{--}0.3$ Jy at 1.4 GHz, where g is the antenna gain at the source's unknown location in the beam normalized to unit amplitude on the beam axis. The other known FRBs typically have peak flux densities an order of magnitude higher, $gS_{1400} \approx 0.2\text{--}2$ Jy. The wide range of flux densities seen at Arecibo, some near the detection threshold, suggests that weaker bursts are also produced, probably at a higher rate. The rate of burst detections is $\sim 3\text{ h}^{-1}$ for bursts with $gS_{1400} \geq 20$ mJy over all observations in which an ALFA beam was within $3.5'$ of the improved position. We note, however, that the bursts appear to cluster in time with some observing sessions showing multiple bright bursts and others showing none.

The observed burst full-widths at half-maximum are $w_{50} = 2.8\text{--}8.7$ ms, which are consistent with the $w_{50} = 1.3\text{--}23.4$ ms widths seen from other FRBs. No clear evidence for scatter broadening was seen in any of the bursts. Bursts 8 and 10 show double-peaked profiles, which has also been seen in FRB 121002 (ref. 7). Furthermore, the morphologies of bursts 8 and 10 evolve smoothly with frequency.

Within our observing band (1.214–1.537 GHz) the burst spectra are remarkably variable. Some are brighter towards higher frequencies, as in the initial discovery, burst 1, while others are brighter towards lower frequencies. The spectra of bursts 8 and 10 are not monotonic. The detections of bursts 6–11 exclusively in beam 0 of the ALFA receiver

¹Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, B-53121 Bonn, Germany. ²Department of Physics and McGill Space Institute, McGill University, 3600 University Street, Montreal, Quebec H3A 2T8, Canada. ³ASTRON, Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA Dwingeloo, The Netherlands. ⁴Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. ⁵Columbia Astrophysics Laboratory, Columbia University, New York, New York 10027, USA. ⁶Department of Astronomy and Space Sciences, Cornell University, Ithaca, New York 14853, USA. ⁷Cornell Center for Advanced Computing, Cornell University, Ithaca, New York 14853, USA. ⁸Square Kilometre Array South Africa, Pinelands, 7405, South Africa. ⁹Department of Physics and Astronomy, Franklin and Marshall College, Lancaster, Pennsylvania 17604-3003, USA. ¹⁰National Research Council, Naval Research Laboratory, 4555 Overlook Avenue SW, Washington DC 20375, USA. ¹¹National Radio Astronomy Observatory, PO Box 2, Green Bank, West Virginia 24944, USA. ¹²Department of Physics and Astronomy, West Virginia University, Morgantown, West Virginia 26506, USA. ¹³National Radio Astronomy Observatory, Charlottesville, West Virginia 22903, USA. ¹⁴Arecibo Observatory, HC3 Box 53995, Arecibo, Puerto Rico 00612, USA. ¹⁵Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada. ¹⁶Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK.

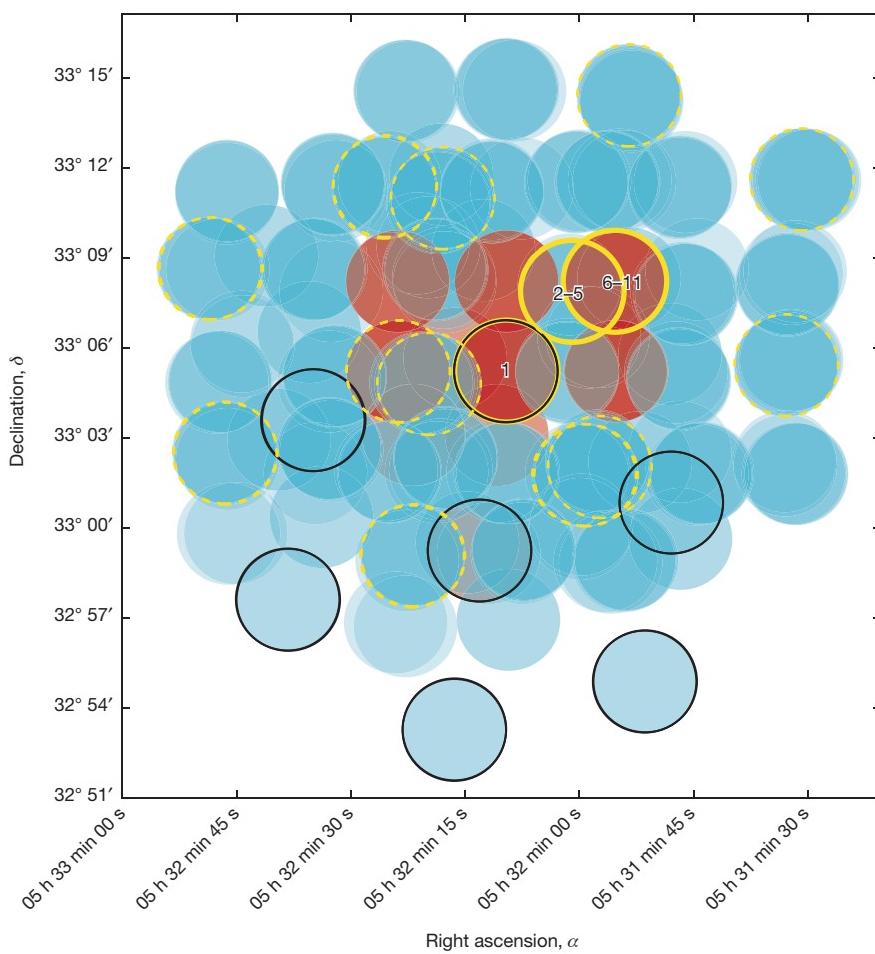


Figure 1 | Discovery and follow-up detections of FRB 121102. For each seven-beam ALFA pointing, the central and outer six beams are shown schematically, in red and blue, respectively (see Extended Data Tables 1 and 2). The circles indicate the $\sim 3.5'$ half-power widths of the beams at 1.4 GHz. Darker shading indicates sky positions with multiple grid observations at roughly the same position. The initial discovery pointing⁴ and second survey observation are outlined in black (these overlap). Beam positions in which bursts were detected are outlined in solid yellow (dashed yellow outlines for the other six beams from the same pointing) and the corresponding burst identifier numbers (Table 1) are given.

(see Extended Data Table 1) means that the bursts must have been detected in the main beam and not in a side-lobe. Although the frequency-dependent shape of the main beam attenuates the bursts' intrinsic spectra at higher frequencies if the source is off-axis⁴, this bias is either not large enough or in the wrong direction to cause the observed spectral variability of bursts 6–11. Given our improved position, burst 1 is consistent with its detection in a side-lobe, which, unlike in the main beam, could have caused attenuation of the spectrum at lower frequencies. This spectral volatility is reflected by the wide range of spectral indices $\alpha \approx -10$ to +14 obtained from fitting a power-law model ($S_\nu \propto \nu^\alpha$, where S_ν is the flux density at radio frequency ν) to burst spectra (Table 1).

There is no evidence for fine-scale diffractive interstellar scintillation, most probably because it is unresolved by our limited spectral resolution. In principle, the spectra could be strongly modulated if the source is multiply imaged by refraction in the interstellar medium¹⁸ or by gravitational lensing. However, the splitting angle between sub-images required to produce spectral structure across our band ($\ll 1$ milliarcsecond) is much smaller than the expected diffraction angle from interstellar plasma scattering. The fine-scaled diffraction structure in the spectrum will therefore wash out the oscillation. Lastly, positive spectral indices could also be explained by free-free absorption at the source¹⁹, but this is ruled out by the large spectral differences

Table 1 | Properties of detected bursts

Burst number	Barycentric peak time (MJD)	Peak flux density (Jy)	Fluence (Jy ms)	Gaussian width (ms)	Spectral index	DM (pc cm ⁻³)
1	56233.282837008	0.04	0.1	3.3 ± 0.3	8.8 ± 1.9	$553 \pm 5 \pm 2$
2	57159.737600835	0.03	0.1	3.8 ± 0.4	2.5 ± 1.7	$560 \pm 2 \pm 2$
3	57159.744223619	0.03	0.1	3.3 ± 0.4	0.9 ± 2.0	$566 \pm 5 \pm 2$
4	57175.693143232	0.04	0.2	4.6 ± 0.3	5.8 ± 1.4	$555 \pm 1 \pm 2$
5	57175.699727826	0.02	0.09	8.7 ± 1.5	1.6 ± 2.5	$558 \pm 6 \pm 4$
6	57175.742576706	0.02	0.06	2.8 ± 0.4		$559 \pm 9 \pm 1$
7	57175.742839344	0.02	0.06	6.1 ± 1.4	-3.7 ± 1.8	
8	57175.743510388	0.14	0.9	6.6 ± 0.1		$556.5 \pm 0.7 \pm 3$
9	57175.745665832	0.05	0.3	6.0 ± 0.3	-10.4 ± 1.1	$557.4 \pm 0.7 \pm 3$
10	57175.747624851	0.05	0.2	8.0 ± 0.5		$558.7 \pm 0.9 \pm 4$
11	57175.748287265	0.31	1.0	3.06 ± 0.04	13.6 ± 0.4	$556.5 \pm 0.1 \pm 1$

Uncertainties are the 68% confidence interval, unless otherwise stated. MJD, modified Julian day.

The barycentric peak time is the arrival time corrected to the Solar System barycentre and referenced to infinite frequency (that is, the time delay due to dispersion is removed).

The peak flux density and the fluence are lower limits because it is assumed that the burst is detected at the centre of the beam (that is, with an assumed gain of 10 KJy^{-1} yielding a system equivalent flux density of 3 Jy). Gaussian widths are the full-width at half-maximum. For the spectral index, bursts 8 and 10 are not well fitted by a power-law model and burst 6 is too corrupted by RFI to include. Quoted errors on DM are, in order, statistical and systematic (see Methods). The DM for burst 7 was too weak and corrupted by RFI to include.

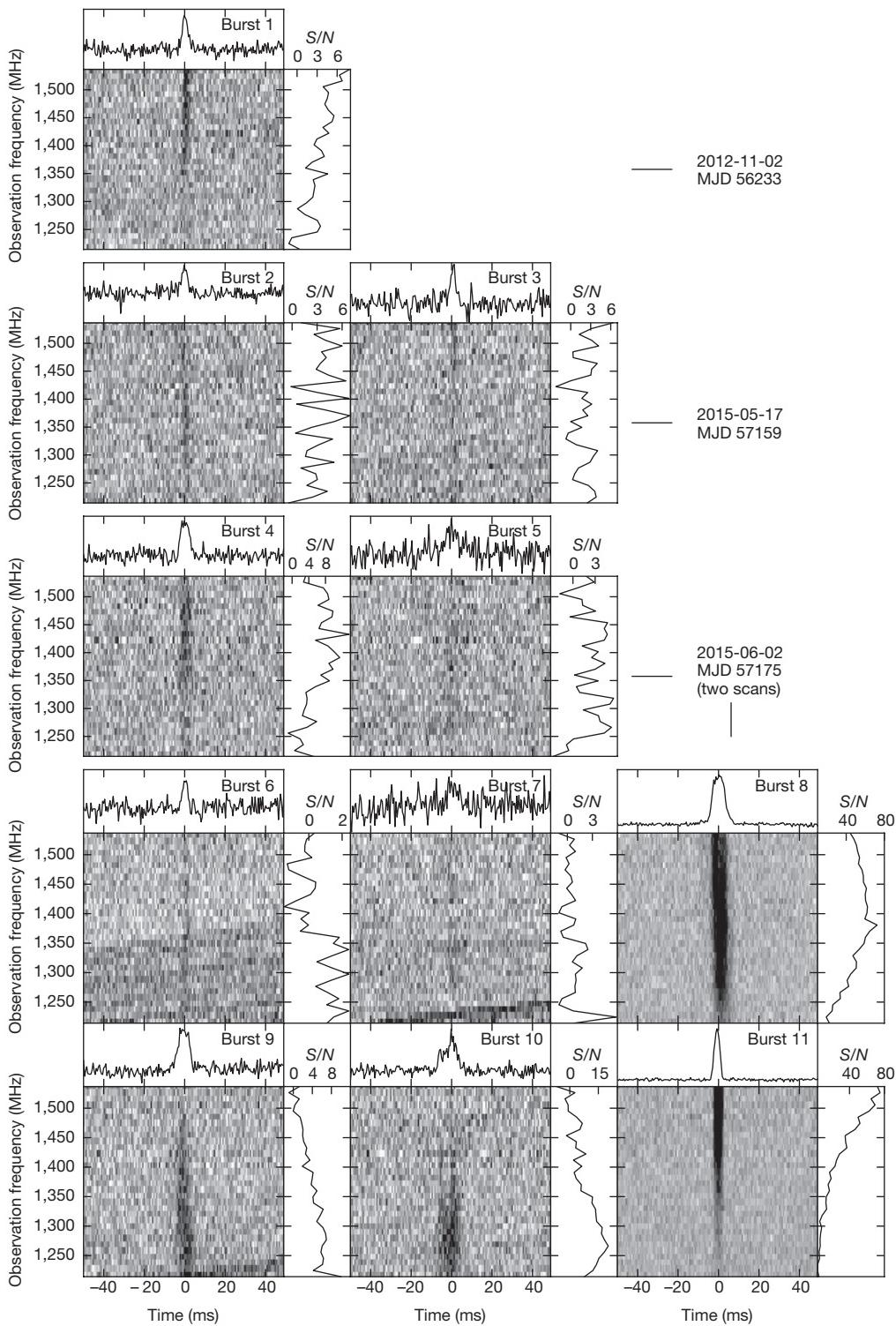


Figure 2 | FRB 121102 burst morphologies and spectra. The central greyscale (linearly scaled) panels show the total intensity versus observing frequency and time, after correcting for dispersion to a value of $DM = 559 \text{ pc cm}^{-3}$. The data are shown with frequency resolution 10 MHz and time resolution 0.524 ms. The diagonal striping at low radio frequencies for bursts 6, 7 and 9 is due to RFI that is unrelated to FRB 121102. The upper sub-panels are burst profiles summed over all frequencies. The band-corrected burst spectra are shown in the right sub-panels. The signal-to-noise ratio (S/N) scales for the spectra are shown on each sub-panel. All panels are arbitrarily and independently scaled.

among bursts. We therefore conclude that the spectral shapes and variations are likely to be predominantly intrinsic to the source.

An analysis of the arrival times of the bursts did not reveal any statistically significant periodicity (see Methods). If the source has a long period (≥ 1 s), then it is probably emitting at a wide range of rotational phases, which is not uncommon for magnetars²⁰, making a convincing period determination difficult. Owing to the small number of detected bursts, we are not sensitive to periodicities much shorter than ~ 100 ms.

Repeat bursts rule out models involving cataclysmic events—such as merging neutron stars²¹ or collapsing super-massive neutron stars¹⁰. Bursts from Galactic flare stars have been proposed as a model for FRBs

with the DM excess originating in the stellar corona²². However, temporal density variations in the corona should produce bursts with varying DMs, which we do not observe. Planets orbiting in a magnetized pulsar wind may produce a millisecond-duration burst once per orbital period²³; however, the observed intra-session separations of our bursts (23–572 s) are too short to correspond to orbital periods. Repeated powerful radiative bursts are associated with magnetars, and indeed giant flares from the latter have been suggested as a FRB source^{12,19,24}. However, no Galactic magnetar has been seen to emit more than a single giant flare in over four decades of monitoring, arguing against a magnetar giant flare origin for FRB 121102. Magnetars have

been observed to exhibit repeating bright radio pulses²⁰, but not yet at the energy scale implied if FRB 121102 is more than several hundred kiloparsecs away.

Giant pulse emission from an extragalactic pulsar remains a plausible model¹¹. The most prominent giant pulses are from the Crab pulsar, which has a large spin-down energy loss rate. Spectral indices calculated from wideband measurements of giant pulses from the Crab pulsar²⁵ have a broad distribution ranging from $\alpha \approx -15$ to $+10$, as well as frequency ‘fringes’—that is, a banded structure to the emission brightness as a function of frequency²⁶. These fringes have characteristic widths of a few hundred megahertz, and we speculate that—given our 322-MHz observing bandwidth—a similar phenomenon could create the spectral variability we have seen in FRB 121102. The double-peaked nature of some FRB 121102 bursts is also possible in the giant pulse model^{7,25}, and the evolution of these burst morphologies with frequency could imply rapid spectral variation between consecutive (sub-)pulses only milliseconds apart.

The low Galactic latitude and relatively small β_{DM} of FRB 121102 compared with other FRBs raises the question of whether it is genuinely extragalactic in origin (see also Methods). However, no H α or H II regions are seen in archival data along the line of sight to FRB 121102, as might be expected for an intervening ionized nebula⁴ that can give $\beta_{\text{DM}} \gg 1$. Furthermore, a detailed multi-wavelength investigation, which searched for a compact nebula in a sky region that includes the refined position presented here, concluded that FRB 121102’s high DM cannot be explained by unmodelled Galactic structure along the line of sight and that FRB 121102 must therefore be extragalactic¹⁹. Conclusively establishing that FRB 121102 is extragalactic will require arcsecond localization and association with a host galaxy. The repeating nature of the bursts facilitates such localization with a radio interferometer.

Although the FRB 121102 bursts share many similarities to the FRBs detected using the Parkes^{1–3,5–7} and Green Bank⁸ telescopes, it is unclear whether FRB 121102 is representative of all FRBs. The ten bursts from FRB 121102 in 2015 were detected near the best-known position in 3 h of observations. In contrast, follow-up observations of the Parkes FRBs, again using the Parkes telescope, range in total time per direction from a few hours⁹ to almost 100 hours¹ and have found no additional bursts. Arecibo’s sensitivity is at least ten times higher, possibly allowing detection of a broader range of the burst-energy distribution of FRBs, and thus increasing the chances of detecting repeated bursts; for example, of the 11 bursts from FRB 121102, Parkes may have been capable of detecting only bursts 8 or 11. More sensitive observations of the Parkes FRBs may therefore show that they also sporadically repeat.

Alternatively, FRB 121102 may be fundamentally different from the FRBs detected at Parkes and Green Bank. As was the case for supernovae and gamma-ray bursts, multiple astrophysical processes may be required to explain the diversity of observational properties of FRBs.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 December 2015; accepted 21 January 2016.

Published online 2 March 2016.

1. Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J. & Crawford, F. A bright millisecond radio burst of extragalactic origin. *Science* **318**, 777–780 (2007).
2. Thornton, D. et al. A population of fast radio bursts at cosmological distances. *Science* **341**, 53–56 (2013).
3. Burke-Spolaor, S. & Bannister, K. W. The Galactic position dependence of fast radio bursts and the discovery of FRB 011025. *Astrophys. J.* **792**, 19 (2014).
4. Spitler, L. G. et al. Fast radio burst discovered in the Arecibo Pulsar ALFA survey. *Astrophys. J.* **790**, 101 (2014).
5. Petroff, E. et al. A real-time fast radio burst: polarization detection and multiwavelength follow-up. *Mon. Not. R. Astron. Soc.* **447**, 246–255 (2015).
6. Ravi, V., Shannon, R. M. & Jameson, A. A fast radio burst in the direction of the Carina dwarf spheroidal galaxy. *Astrophys. J.* **799**, L5 (2015).
7. Champion, D. J. et al. Five new fast radio bursts from the HTRU high latitude survey: first evidence for two-component bursts. Preprint at <http://arxiv.org/abs/1511.07746> (2015).

8. Masui, K. et al. Dense magnetized plasma associated with a fast radio burst. *Nature* **528**, 523–525 (2015).
9. Petroff, E. et al. A survey of FRB fields: limits on repeatability. *Mon. Not. R. Astron. Soc.* **454**, 457–462 (2015).
10. Falcke, H. & Rezzolla, L. Fast radio bursts: the last sign of supramassive neutron stars. *Astron. Astrophys.* **562**, A137 (2014).
11. Cordes, J. M. & Wasserman, I. Supergiant pulses from extragalactic neutron stars. *Mon. Not. R. Astron. Soc.* **457**, 232–257 (2016).
12. Pen, U.-L. & Connor, L. Local circumnuclear magnetar solution to extragalactic fast radio bursts. *Astrophys. J.* **807**, 179 (2015).
13. Cordes, J. M. et al. Arecibo pulsar survey using ALFA. I. Survey strategy and first discoveries. *Astrophys. J.* **637**, 446–455 (2006).
14. Lazarus, P. et al. Arecibo pulsar survey using ALFA. IV. Mock spectrometer data analysis, survey sensitivity, and the discovery of 40 pulsars. *Astrophys. J.* **812**, 81 (2015).
15. Cordes, J. M. & Lazio, T. J. W. NE2001.I. A new model for the galactic distribution of free electrons and its fluctuations. Preprint at <http://arxiv.org/abs/astro-ph/0207156> (2002).
16. Ransom, S. M. *New Search Techniques for Binary Pulsars*. http://www.cv.nrao.edu/~sransom/ransom_thesis_2001.pdf, PhD thesis, Harvard Univ. (2001).
17. Karakao-Argaman, C. et al. Discovery and follow-up of rotating radio transients with the Green Bank and LOFAR telescopes. *Astrophys. J.* **809**, 67 (2015).
18. Wolszczan, A. & Cordes, J. M. Interstellar interferometry of the pulsar PSR 1237+25. *Astrophys. J.* **320**, L35–L39 (1987).
19. Kulkarni, S. R., Ofek, E. O. & Neill, J. D. The Arecibo fast radio burst: dense circum-burst medium. Preprint at <http://arxiv.org/abs/1511.09137> (2015).
20. Camilo, F. et al. Transient pulsed radio emission from a magnetar. *Nature* **442**, 892–895 (2006).
21. Hansen, B. M. S. & Lyutikov, M. Radio and X-ray signatures of merging neutron stars. *Mon. Not. R. Astron. Soc.* **322**, 695–701 (2001).
22. Loeb, A., Shvartzvald, Y. & Maoz, D. Fast radio bursts may originate from nearby flaring stars. *Mon. Not. R. Astron. Soc.* **439**, L46–L50 (2014).
23. Mottez, F. & Zarka, P. Radio emissions from pulsar companions: a refutable explanation for galactic transients and fast radio bursts. *Astron. Astrophys.* **569**, A86 (2014).
24. Lyubarsky, Y. A model for fast extragalactic radio bursts. *Mon. Not. R. Astron. Soc.* **442**, L9–L13 (2014).
25. Karuppusamy, R., Stappers, B. W. & van Straten, W. Giant pulses from the Crab pulsar. A wide-band study. *Astron. Astrophys.* **515**, A36 (2010).
26. Hankins, T. H. & Eilek, J. A. Radio emission signatures in the Crab pulsar. *Astrophys. J.* **670**, 693–701 (2007).

Acknowledgements We thank the staff of the Arecibo Observatory, and in particular A. Venkataraman, H. Hernandez, P. Perillat and J. Schmelz, for their continued support and dedication to enabling observations like those presented here. We also thank our commensal observing partners from the Arecibo ‘Zone of Avoidance’ team, in particular T. McIntyre and T. Henning. We thank M. Kramer for suggestions. The Arecibo Observatory is operated by SRI International under a cooperative agreement with the National Science Foundation (AST-1100968), and in alliance with Ana G. Méndez-Universidad Metropolitana, and the Universities Space Research Association. These data were processed using the McGill University High Performance Computing Centre operated by Compute Canada and Calcul Québec. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The research leading to these results has received funding from the European Research Council (ERC) under the European Union’s Seventh Framework Programme (FP7/2007–2013). L.G.S., P.C.C.F. and P.L. gratefully acknowledge financial support from the ERC Starting Grant BEACON under contract number 279702. J.W.T.H. is an NWO Vidi Fellow and gratefully acknowledges funding for this work from ERC Starting Grant DRAGNET under contract number 337062. Work at Cornell (J.M.C., S.C., A.B.) was supported by NSF grants AST-1104617 and AST-1008213. V.M.K. holds the Lorne Trottier Chair in Astrophysics and Cosmology and a Canadian Research Chair in Observational Astrophysics and received additional support from NSERC via a Discovery Grant and Accelerator Supplement, by FQRNT via the Centre de Recherche Astrophysique de Québec, and by the Canadian Institute for Advanced Research. J.V.L. acknowledges funding for this research from an ERC Consolidator Grant under contract number 617199. Pulsar research at UBC is supported by an NSERC Discovery Grant and by the Canadian Institute for Advanced Research.

Author Contributions L.G.S. and J.W.T.H. led the design and execution of the observing campaign described here. P.S. performed the analysis that discovered the radio bursts. More detailed analysis of the signal properties was done by L.G.S., P.S., S.M.R., M.A.M., J.W.T.H., S.C. and J.M.C. L.G.S., J.W.T.H., P.S. and V.M.K. led the writing of the manuscript. All authors contributed substantially to the interpretation of the analysis results and to the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W.T.H. (j.w.t.hessels@uva.nl).

METHODS

Positional gridding strategy and burst localization. Extended Data Table 1 summarizes all the observations made in both the initial 2013 follow-up and 2015 May/June observations (project code p2886). The ALFA beam 0 pointing positions in J2000 equatorial coordinates are summarized in Extended Data Table 2. In the p2030/p2886 observations, the major axis of the ALFA receiver was rotated 19°/90° with respect to North⁴.

In 2015 May/June, we searched for additional bursts from FRB 121102 using a grid of six pointings using the seven-beam ALFA receiver to cover a generous ∼9' radius around the discovery beam position and side-lobe. The ALFA receiver was aligned East–West to optimize the sky coverage for this specific purpose. The centre beams of the six grid pointings are shown in red in Fig. 1, and the six outer ALFA beams are shown in blue. Each grid pointing position was observed at least four times for ∼1,000 s. The beam positions of the discovery observation and 2013 follow-up gridding⁴ with ALFA (in that case rotated 19° with respect to North) are also indicated using the same colour scheme. The outer six ALFA beams in the multiple grid observations are only at roughly the same position because the projection of the ALFA beams on the sky depends on the position of the telescope feed with respect to the primary reflecting dish, and these do not overlap perfectly between independent observations.

Two bursts on May 17 (bursts 2 and 3) and two on June 2 (bursts 4 and 5) were detected at a single grid position: FRBGRID2b in ALFA beam 6, which had positions of $\alpha = 05^{\circ}32\text{ min}01\text{s}$, $\delta = +33^{\circ}07'\text{56''}$ and $\alpha = 05^{\circ}32\text{ min}01\text{s}$, $\delta = +33^{\circ}07'\text{53''}$ (J2000) at the two epochs—that is, only a few arcseconds apart. Six more bursts (bursts 6–11) were detected on June 2 at neighbouring grid position, FRBGRID6b in ALFA beam 0, ∼1.3' away at $\alpha = 05^{\circ}31\text{ min}55\text{s}$, $\delta = +33^{\circ}08'\text{13''}$. In all cases bursts were detected in only one beam of the seven-beam ALFA receiver at any given time. This shows that the bursts must originate beyond Arecibo's Fresnel length of ∼100 km (ref. 19).

The intermittency of FRB 121102 makes accurate localization more challenging. Nonetheless, the detection in adjacent grid positions is informative, and to refine the position of FRB 121102, we simply take the average position between FRBGRID2b ALFA beam 6 and FRBGRID6b ALFA beam 0, which gives: $\alpha = 05^{\circ}31\text{ min}58\text{s}$, $\delta = +33^{\circ}08'\text{04''}$ (J2000) and, equivalently, Galactic longitude and latitude $l = 174.89^{\circ}$, $b = -0.23^{\circ}$. The approximate uncertainty radius of ∼3' is based on the amount of overlap between the two detection beam positions and the ALFA beam width at half power, which is ∼3.5'. The distance from the initially reported burst 1 position is 3.7', consistent with the interpretation that this burst was detected in a side-lobe. Although FRB 121102 bursts have been detected in beams with different central sky positions, all detections are consistent with a well defined sky position when one considers the imprint of the ALFA gain pattern on the sky during each observation⁴.

Galactic versus extragalactic interpretation. Noteworthy is the fact that FRB 121102 lies directly in the Galactic plane, whereas the other claimed FRBs lie predominantly at high Galactic latitudes. The PALFA survey is only searching in the Galactic plane, however, and no comparable FRB survey at 1.4 GHz with Arecibo has been done at high Galactic latitudes. Therefore, this difference may simply be a consequence of where Arecibo has most deeply searched for FRBs and does not necessarily suggest that FRB 121102 is of Galactic origin. Furthermore, FRB 121102 was found in the Galactic anti-centre region of the PALFA survey, whereas searches in the inner-Galaxy region have thus far found no FRBs¹⁴. This may be because the Galactic foregrounds in the anti-centre region are comparatively low, so the deleterious effects of DM smearing and scattering, which may reduce our sensitivity to FRBs, are less important in the outer Galaxy than the inner Galaxy.

The low Galactic latitude of FRB 121102 also contributes to its low DM excess factor $\beta_{\text{DM}} \approx 3$ compared to the $\beta_{\text{DM}} \approx 1.2\text{--}40$ range seen for the other 15 FRBs in the literature. Only FRB 010621 (ref. 27), with $\beta_{\text{DM}} \approx 1.2$, has a lower β_{DM} than FRB 121102, and it has been proposed to be Galactic²⁸. We note, however, that six of 16 FRBs have DMs comparable to or lower than FRB 121102. Furthermore, its total Galactic DM excess $\text{DM}_{\text{FRB}} - \text{DM}_{\text{max}}^{\text{Gal}} \approx 370 \text{ pc cm}^{-3}$ is larger than that of the first-discovered FRB¹. Lastly, within a generous 20-degree radius of FRB 121102, the highest-DM pulsar known is the millisecond pulsar PSR J0557+1550 (ref. 29; also a PALFA discovery), which has $\text{DM} = 103 \text{ pc cm}^{-3}$ and $\beta_{\text{DM}} = 0.6$, as well as the highest DM-inferred distance¹⁵ of any pulsar in this region, $d = 5.7 \text{ kpc}$. FRB 121102's DM is clearly anomalous, even when compared to this distant Galactic anti-centre pulsar. At an angular offset of 38°, we note the existence of PSR J0248+6021, with $\text{DM} = 370 \text{ pc cm}^{-3}$ and $\beta_{\text{DM}} = 1.8$. Although the DM of this young, 217-ms pulsar is in excess of the maximum Galactic contribution in the NE2001 model¹⁵, this can be explained by its location within the dense, giant H II region W5 in the Perseus arm³⁰ at a distance of 2 kpc. A similar association for FRB 121102 has been sought to explain its $\beta_{\text{DM}} \approx 3$, but multi-wavelength investigations have as yet found no unmodelled Galactic structure^{4,19}. In summary,

FRB 121102's comparatively low β_{DM} does not strongly distinguish it from other FRBs, or necessarily suggest it is more likely to be Galactic.

Observations and search processing. Here we provide a brief description of the Arecibo Mock spectrometer data and search pipeline¹⁴ used for our follow-up observations of FRB 121102. The 1.4-GHz data were recorded with the Mock spectrometers, which cover the full ALFA receiver bandwidth in two subbands. Each 172-MHz subband was sampled with 16 bits, a time resolution of $65.5 \mu\text{s}$, and frequency resolution of 0.34 MHz in 512 channels. The data were later converted to 4-bit samples to reduce the data storage requirements. Before processing, the two subbands were combined into a single band of 322 MHz (accounting for frequency overlap between the two subbands), which was centred at 1,375 MHz and spans 1,214.3–1,536.7 MHz.

We used the PALFA PRESTO-based¹⁶ search pipeline¹⁴ to search for astrophysical signals in the frequency and time domains. These data were processed using the McGill University High Performance Computing Centre operated by Compute Canada and Calcul Québec. The presence of RFI can have a detrimental effect on our ability to detect bursts. We therefore applied PRESTO's *rffind* software tool to identify contaminated frequency channels and time blocks. Flagged channels and time blocks were masked in subsequent analyses. Time blocks contaminated by RFI are identified using data that are not corrected for dispersive delay (that is, $\text{DM} = 0 \text{ pc cm}^{-3}$), in order to avoid removing astrophysical signals. The data were corrected for dispersion using 7,292 trial DMs in the range $0\text{--}9,866.4 \text{ pc cm}^{-3}$, generating a time series at each trial. We performed Fourier analyses of all the time series to look for periodic signals using PRESTO's *accelsearch* software tool and detected no significant signal of a plausible astrophysical origin.

We searched for single pulses in each dispersion-corrected time series by convolving a template bank of boxcar functions with widths ranging from 0.13 ms to 100 ms. This optimizes the detection of pulses with durations longer than the native sample time of the data. Single-pulse events at each DM were identified by applying a signal-to-noise ratio (*S/N*) threshold of 5.

These single-pulse events were grouped and ranked using the RRATtrap sifting algorithm¹⁷. An astrophysical pulse is detected with maximum *S/N* at the signal's true DM and is detected with decreasing *S/N* at nearby trial DMs. This is not generally the case for RFI, whose *S/N* does not typically peak at a non-zero trial DM. The RRATtrap algorithm ranks candidates based on this DM behaviour and candidate plots are produced for highly ranked single-pulse groups. These plots display the *S/N* of the pulse as a function of DM and time as well as an image of the signal as a function of time and observing frequency (for example, Fig. 2). The resulting plots were inspected for astrophysical signals, and pulses were found at a DM of $\sim 559 \text{ pc cm}^{-3}$ at a sky position consistent with the discovery position of FRB 121102 (ref. 4). It is possible that the analysed data contain weaker bursts, which cannot be reliably identified because their *S/N* is too low to distinguish them from RFI or statistical noise. If, in the future, the bursts are shown to have an underlying periodicity, then this would enable a deeper search for weak bursts.

Timing analysis of burst arrival times. Using several approaches, we searched for an underlying periodicity matching the arrival times of the eight bursts detected in the 2015 June 2 observing session. There are no significant periodicities detected through a standard fast Fourier transform of the time series. We then carried out a similar analysis to that routinely used to detect periodicities in sporadically emitting radio pulsars³¹. In this analysis, we calculate differences between all of the burst arrival times and search for the greatest common denominator of these differences. We found several periods, not harmonically related, that fitted different subsets of bursts within a tolerance of 1% of the trial period, but none that fitted all of the bursts. We subsequently calculated residuals for the times-of-arrival for the eight bursts detected on 2015 June 2 for a range of trial periods using the pulsar timing packages TEMPO and PINT (see 'Code availability' section). We found that some of the periods returned by the differencing algorithm also resulted in residuals with root-mean-square value of less than 1% of the trial period. However, there were many non-harmonically related candidate periods resulting in residuals of a comparable root-mean-square value. Furthermore, given the number of trials necessary for this search, none of these trial periods was statistically significant. In addition, owing to the small number of detected bursts, and the widths of the pulses, we were not sensitive to periodicities much shorter than ∼100 ms because our tolerance for a period match (or acceptable root-mean-square value) becomes a large fraction of the period and there are many possible fits. The 16-day gap between the 2015 May and June detections precluded us from including the May bursts in any search for periodicity in the single pulses.

Spectral fitting. To produce the spectra shown in the right panels of Fig. 2, we corrected each spectrum for the bandpass of the receiver. We estimated the bandpass by taking the average of the raw data samples for each frequency channel. We then median-filtered that average bandpass with a width of 20 channels to remove the effects of narrow-band RFI and divided the observed spectrum of each burst by this median-filtered bandpass. The band-corrected burst spectra shown in the

right sub-panels of Fig. 2 are still somewhat contaminated by RFI, however. The bottom and top ten channels (3.4 MHz) of the band were ignored owing to roll-off in the receiver response.

To characterize the bandpass-corrected spectrum of each burst, we applied a power-law model using least-squares fitting. The power-law model is described by $S_\nu \propto \nu^\alpha$, where S_ν is the flux density in a frequency channel, ν is the observing frequency, and α is the spectral index. These measured spectral indices and their uncertainties are shown in Table 1. We do not include a spectral index value for burst 6 because of the RFI in the lower half of the band. For bursts 7 and 9, we exclude data below 1,250 MHz, because of RFI contamination. For bursts 8 and 10, the power-law model was not a good descriptor, and therefore no value is reported in Table 1.

We verified this technique by applying the bandpass correction to PALFA data of pulsar B1900+01. The measured spectral index was calculated for ten bright single pulses, and the values are consistent with the published value.

Measurement of DM and dispersion index. We measured the DM for 10 of the 11 bursts and additionally the dispersion index ξ (from the dispersive delay $\Delta t \propto \nu^{-\xi}$) for the brightest two. The DM and the dispersion index were calculated with a least-squares routine using the SIMPLEX and MIGRAD functions from the CERN MINUIT package (<http://www.cern.ch/minuit>). The user specifies the assumed form of the intrinsic pulse shape, which is then convolved with the appropriate DM smearing factor. For these fits a boxcar pulse template was used.

Subbanded pulse profiles for each burst were generated by averaging blocks of frequency channels. The number of subbands generated depended on the S/N of the burst to ensure that there was sufficient S/N in each subband for the fit to converge. Subbands with no signal were excluded from the fit. Furthermore, the data were binned in time to further increase the S/N and reduce the effects of frequency-dependent flux evolution. As the true intrinsic pulse width is not known, each burst was fitted with a range of boxcar widths. The parameters corresponding to the input template yielding the cleanest residuals are reported.

The DM value was fitted keeping the DM index fixed at 2.0. We note that burst 7 was too weak and corrupted by RFI to obtain reasonable fits. Additionally, for the brightest two bursts (8 and 11), we also did a joint fit of DM and dispersion index. The resulting dispersion index fits were 2.00 ± 0.02 and 1.999 ± 0.002 for bursts 8 and 11, respectively. These values are as expected for radio waves travelling through a cold, ionized medium.

Frequency-dependent pulse profile evolution introduces systematic biases into the times of arrival in each subband. These biases in turn bias the DM determination. These systematics cannot be mitigated without an accurate model for the underlying burst shape versus frequency, which is not available in this case, and is further complicated by the fact that the burst morphology also changes randomly from burst to burst. We estimated the systematic uncertainty by considering what DM value would produce a delay across our observing band that is comparable to half the burst width in each case.

Table 1 presents the results of the fits with the statistical and systematic uncertainties both quoted. The DM estimates do not include barycentric corrections (of the order of $0.01\text{--}0.1 \text{ pc cm}^{-3}$). Although FRB 121102 is close to the ecliptic, the angular separation from the Sun was always much larger than 10° , and any annual contribution to the DM from the solar wind was small ($<10^{-3} \text{ pc cm}^{-3}$)^{32,33}. These effects are, therefore, much smaller than the aforementioned systematics in modelling the DMs of the bursts.

The $\pm 1\sigma$ range of DMs for the ten new bursts is $558.1 \pm 3.3 \text{ pc cm}^{-3}$, consistent with the discovery value⁴, $557.4 \pm 2.0 \text{ pc cm}^{-3}$. The DM values and dispersion indices reported here and previously⁴ were calculated using different methods. These two approaches fitted for different free parameters, so different covariances between parameters may result in slightly different values. Also, different time and frequency resolutions were used. Nonetheless, the burst 1 parameters quoted here and previously⁴ are consistent within the uncertainties. The consistency of the DM values is conclusive evidence that a single source is responsible for the events.

The role of interstellar scattering. Some FRBs have shown clear evidence for multi-path propagation from scattering by the intervening interstellar or extragalactic material along the line of sight^{2,6–8}. However, the burst profiles from FRB 121102 show no obvious evidence for asymmetry from multi-path propagation. An upper bound⁴ on the pulse broadening time from burst 1 is 1.5 ms at 1.5 GHz. Using the NE2001 model for a source far outside the Galaxy, the expected pulse broadening is $\sim 20 \mu\text{s} \times \nu^{-4.4}$ with ν in gigahertz, an order of magnitude smaller than the $\sim 2\text{-ms}$ pulse widths and $\sim 0.7\text{-ms}$ intra-channel dispersion smearing. The features of the spectra cannot be explained by diffractive interstellar scintillations; the predicted scintillation bandwidth for FRB 121102 is $\sim 50 \text{ kHz}$ at 1.5 GHz, which is unresolved by the 0.34-MHz frequency channels of our data. We would, therefore, also not expect to observe diffractive interstellar scintillation in our bursts. Additional scattering occurring in a host galaxy and the intergalactic medium is at a level below our ability to detect. However, observations at frequencies below 1.5 GHz may reveal pulse broadening that is not substantially smaller than the upper bound if we use as a guide the observed pulse broadening from other FRBs^{2,6–8}. Future observations that quantify diffractive interstellar scintillations can provide constraints on the location of extragalactic scattering plasma relative to the source, as demonstrated for FRB 110523 (ref. 8).

The upper bound on pulse broadening for FRB 121102 implies that the apparent, scattered source size for radio waves incident on the Milky Way's interstellar medium is small enough that refractive interstellar scintillation (RISS) from the interstellar medium is expected. For the line of sight to FRB 121102, we use the NE2001 model to estimate an effective scattering-screen distance of $\sim 2 \text{ kpc}$ from Earth and a scattering diameter of 6 milliarcseconds. The implied length scale for phase-front curvature is then $l_{\text{RISS}} \approx 2 \text{ kpc} \times 6 \text{ milliarcseconds} = 12 \text{ AU}$. For an effective, nominal velocity, $V_{\text{eff}} = 100 \times V_{100} \text{ km s}^{-1}$, the expected RISS timescale is $\Delta t_{\text{RISS}} = l_{\text{RISS}} / V_{\text{eff}} = 215 V_{100}^{-1} \nu_{1\text{GHz}}^{-2.2} \text{ days}$. At 1.5 GHz and with an effective velocity due to Galactic rotation of about 200 km s^{-1} in the direction of FRB 121102, RISS timescales of 20–40 days are expected. Modulation from RISS can be several tens of per cent³⁴. This level of modulation could play a part in the detections of bursts in 2015 mid-May and 2015 June and their absence in 2015 early-May and at other epochs. However, the Solar System and the ionized medium have the same Galactic rotation, so the effective velocity could be smaller than 100 km s^{-1} , leading to longer RISS timescales.

Data availability. The beam positions used in Fig. 1 and the data of the bursts used to generate Fig. 2 are provided as Source Data files (available online with the figures).

Code availability. The code used to analyse the data are available at the following sites: PRESTO (<https://github.com/scottransom/presto>), RRATtrap (<https://github.com/ckarako/RRATtrap>), TEMPO (<http://tempo.sourceforge.net/>), and PINT (<http://github.com/nanograv/PINT>).

27. Keane, E. F., Stappers, B. W., Kramer, M. & Lyne, A. G. On the origin of a highly dispersed coherent radio burst. *Mon. Not. R. Astron. Soc.* **425**, L71–L75 (2012).
28. Bannister, K. W. & Madsen, G. J. A Galactic origin for the fast radio burst FRB010621. *Mon. Not. R. Astron. Soc.* **440**, 353–358 (2014).
29. Scholz, P. et al. Timing of five millisecond pulsars discovered in the PALFA survey. *Astrophys. J.* **805**, 85 (2015).
30. Theureau, G. et al. PSRs J0248+6021 and J2240+5832: young pulsars in the northern Galactic plane. Discovery, timing, and gamma-ray observations. *Astron. Astrophys.* **525**, A94 (2011).
31. McLaughlin, M. A. et al. Transient radio bursts from rotating neutron stars. *Nature* **439**, 817–820 (2006).
32. You, X. P., Coles, W. A., Hobbs, G. B. & Manchester, R. N. Measurement of the electron density and magnetic field of the solar wind using millisecond pulsars. *Mon. Not. R. Astron. Soc.* **422**, 1160–1165 (2012).
33. The NANOGrav Collaboration. The NANOGrav nine-year data set: observations, arrival time measurements, and analysis of 37 millisecond pulsars. *Astrophys. J.* **813**, 65 (2015).
34. Rickett, B. J. Radio propagation through the turbulent interstellar plasma. *Annu. Rev. Astron. Astrophys.* **28**, 561–605 (1990).

Extended Data Table 1 | Arecibo FRB 121102 discovery and follow-up observations

UTC Date	Project	Position	Receiver	Frequency (GHz)	Backend	Dwell time (s)	# Bursts
Survey discovery observations presented in Spitler et al. (2014)							
2012-11-02	p2030	FRBDISC	ALFA	1.4	Mocks	200	1 (Beam4)
2012-11-04	p2030	FRBDISC	ALFA	1.4	Mocks	200	0
Follow-up observations presented in Spitler et al. (2014)							
2013-12-09	p2886	FRBGRID1a	ALFA	1.4	Mocks	2700	0
2013-12-09	p2886	FRBGRID2a	ALFA	1.4	Mocks	970	0
2013-12-09	p2886	FRBGRID2a	ALFA	1.4	Mocks	1830	0
2013-12-10	p2886	FRBGRID3a	ALFA	1.4	Mocks	2700	0
2013-12-10	p2886	FRBDISC	327-MHz	0.327	PUPPI	2385	0
Follow-up observations presented here for the first time							
2015-05-02	p2886	FRBDISC	L-wide	1.4	PUPPI	7200	0
2015-05-03	p2886	FRBGRID1b	ALFA	1.4	Mocks	1502	0
2015-05-03	p2886	FRBGRID2b	ALFA	1.4	Mocks	1502	0
2015-05-03	p2886	FRBGRID3b	ALFA	1.4	Mocks	343	0
2015-05-03	p2886	FRBGRID3b	ALFA	1.4	Mocks	1502	0
2015-05-03	p2886	FRBGRID1b	ALFA	1.4	Mocks	921	0
2015-05-05	p2886	FRBGRID1b	ALFA	1.4	Mocks	1002	0
2015-05-05	p2886	FRBGRID2b	ALFA	1.4	Mocks	1002	0
2015-05-05	p2886	FRBGRID3b	ALFA	1.4	Mocks	1002	0
2015-05-05	p2886	FRBGRID4b	ALFA	1.4	Mocks	1002	0
2015-05-05	p2886	FRBGRID5b	ALFA	1.4	Mocks	1002	0
2015-05-05	p2886	FRBGRID6b	ALFA	1.4	Mocks	1002	0
2015-05-09	p2886	FRBGRID1b	ALFA	1.4	Mocks	1002	0
2015-05-09	p2886	FRBGRID2b	ALFA	1.4	Mocks	1002	0
2015-05-09	p2886	FRBGRID3b	ALFA	1.4	Mocks	1002	0
2015-05-09	p2886	FRBGRID4b	ALFA	1.4	Mocks	1002	0
2015-05-09	p2886	FRBGRID5b	ALFA	1.4	Mocks	1002	0
2015-05-09	p2886	FRBGRID6b	ALFA	1.4	Mocks	1002	0
2015-05-09	p2886	FRBGRID6b	ALFA	1.4	Mocks	425	0
2015-05-17	p2886	FRBGRID1b	ALFA	1.4	Mocks	1002	0
2015-05-17	p2886	FRBGRID2b	ALFA	1.4	Mocks	1002	2 (Beam6)
2015-05-17	p2886	FRBGRID3b	ALFA	1.4	Mocks	1002	0
2015-05-17	p2886	FRBGRID4b	ALFA	1.4	Mocks	707	0
2015-05-17	p2886	FRBGRID4b	ALFA	1.4	Mocks	391	0
2015-05-17	p2886	FRBGRID5b	ALFA	1.4	Mocks	1002	0
2015-05-17	p2886	FRBGRID6b	ALFA	1.4	Mocks	1002	0
2015-06-02	p2886	FRBGRID1b	ALFA	1.4	Mocks	1002	0
2015-06-02	p2886	FRBGRID2b	ALFA	1.4	Mocks	1002	2 (Beam6)
2015-06-02	p2886	FRBGRID3b	ALFA	1.4	Mocks	1002	0
2015-06-02	p2886	FRBGRID4b	ALFA	1.4	Mocks	1002	0
2015-06-02	p2886	FRBGRID5b	ALFA	1.4	Mocks	1002	0
2015-06-02	p2886	FRBGRID6b	ALFA	1.4	Mocks	1002	6 (Beam0)
2015-06-02	p2886	FRBGRID6b	ALFA	1.4	Mocks	300	0

The observing setup of these observations is described in Methods. Data from ref. 4 and this study.

Extended Data Table 2 | FRB 121102 gridding positions

Grid ID	Right Ascension	Declination
FRBDISC	05 ^h 32 ^m 09 ^s	+33 ^d 05 ^m 13 ^s
FRBGRID1a	05 ^h 32 ^m 16 ^s	+33 ^d 05 ^m 39 ^s
FRBGRID2a	05 ^h 32 ^m 22 ^s	+33 ^d 03 ^m 06 ^s
FRBGRID3a	05 ^h 32 ^m 11 ^s	+33 ^d 03 ^m 06 ^s
FRBGRID1b	05 ^h 32 ^m 10 ^s	+33 ^d 05 ^m 13 ^s
FRBGRID2b	05 ^h 32 ^m 24 ^s	+33 ^d 05 ^m 13 ^s
FRBGRID3b	05 ^h 31 ^m 55 ^s	+33 ^d 05 ^m 13 ^s
FRBGRID4b	05 ^h 32 ^m 10 ^s	+33 ^d 08 ^m 13 ^s
FRBGRID5b	05 ^h 32 ^m 24 ^s	+33 ^d 08 ^m 13 ^s
FRBGRID6b	05 ^h 31 ^m 55 ^s	+33 ^d 08 ^m 13 ^s

The J2000 ALFA beam 0 positions are listed.

Exponential protection of zero modes in Majorana islands

S. M. Albrecht^{1*}, A. P. Higginbotham^{1,2*}, M. Madsen¹, F. Kuemmeth¹, T. S. Jespersen¹, J. Nygård¹, P. Krogstrup¹ & C. M. Marcus¹

Majorana zero modes are quasiparticle excitations in condensed matter systems that have been proposed as building blocks of fault-tolerant quantum computers¹. They are expected to exhibit non-Abelian particle statistics, in contrast to the usual statistics of fermions and bosons, enabling quantum operations to be performed by braiding isolated modes around one another^{1,2}. Quantum braiding operations are topologically protected insofar as these modes are pinned near zero energy, with the departure from zero expected to be exponentially small as the modes become spatially separated^{3,4}. Following theoretical proposals^{5,6}, several experiments have identified signatures of Majorana modes in nanowires with proximity-induced superconductivity^{7–11} and atomic chains¹², with small amounts of mode splitting potentially explained by hybridization of Majorana modes^{13–15}. Here, we use Coulomb-blockade spectroscopy in an InAs nanowire segment with epitaxial aluminium, which forms a proximity-induced superconducting Coulomb island (a ‘Majorana island’) that is isolated from normal-metal leads by tunnel barriers, to measure the splitting of near-zero-energy Majorana modes. We observe exponential suppression of energy splitting with increasing wire length. For short devices of a few hundred nanometres, sub-gap state energies oscillate as the magnetic field is varied, as is expected for hybridized Majorana modes. Splitting decreases by a factor of about ten for each half a micrometre of increased wire length. For devices longer than about one micrometre, transport in strong magnetic fields occurs through a zero-energy state that is energetically isolated from a continuum, yielding uniformly spaced Coulomb-blockade conductance peaks, consistent with teleportation via Majorana modes^{16,17}. Our results help to explain the trivial-to-topological transition in finite systems and to quantify the scaling of topological protection with end-mode separation.

The set of structures we investigate consists of InAs nanowires grown by molecular beam epitaxy in the [0001] wurtzite direction with an epitaxial aluminium (Al) shell on two facets of the hexagonal cross-section¹⁸. The Al shell was removed except in a small segment of length L and isolated from normal metal (titanium/gold) leads by electrostatic gate-controlled barriers (Fig. 1a). The charging energies E_C of the measured devices range from greater than to less than the superconducting gap of Al (approximately 0.2 meV). The thinness of the Al shell (8–10 nm on the two facets) results in a large critical field B_c before superconductivity is destroyed: for fields along the wire axis, $B_{c,\parallel} \approx 1$ T; out of the plane of the substrate, but roughly in the plane of the two Al-covered facets, $B_{c,\perp} \approx 700$ mT (Fig. 1b). The very high critical fields that are achieved make these wires a suitable platform for investigating topological superconductivity¹⁸.

Five devices over a range of Al shell lengths $L \approx 0.3\text{--}1.5\ \mu\text{m}$ were measured (see Methods for device layouts). Charge occupation and tunnel coupling to the leads were tuned via electrostatic gates. Differential conductance g in the Coulomb-blockade regime (high-resistance

barriers) was measured using standard a.c. lock-in techniques in a dilution refrigerator (electron temperature of about 50 mK).

Figure 1c shows g as a function of gate voltage V_G and source-drain bias V_{SD} . For the $L = 790$ nm device, the zero-field data (Fig. 1c, top) show a series of evenly spaced Coulomb diamonds with a characteristic negative-differential conductance (NDC) region at higher bias. NDC is known from metallic superconductor islands^{19,20} and has recently been reported in a proximitized semiconductor device similar to those investigated here²¹. The zero-magnetic-field diamonds reflect charge transport via Cooper pairs, with gate-voltage period proportional to $2e$, the charge of a Cooper pair. At moderate magnetic fields (Fig. 1c, middle), the large diamonds shrink and a second set of diamonds appears, yielding even-odd spacing of Coulomb-blockade zero-bias conductance peaks²², as seen in the bottom panel of Fig. 1d. At larger magnetic fields (Fig. 1c, bottom), Coulomb diamonds are again periodic, but have precisely half the spacing of the zero-field diamonds, corresponding to $1e$ periodicity. At this field NDC is absent, and resonant structure is visible within each diamond, indicating transport through discrete resonances at low bias and a continuum at high bias (see magnification in Fig. 1c). Coulomb-blockade conductance peaks at high magnetic field (see Fig. 1d for zero-bias cross-sections) with regular $1e$ periodicity (half the zero-field spacing) accompanied by a discrete sub-gap spectrum are a proposed signature of electron teleportation by Majorana end states^{16,17}. We designate the ungrounded tunnelling device in this high-field regime as a ‘Majorana island’, where a sub-gap state near zero energy, energetically isolated from a continuum, leads to $1e$ -periodic Coulomb-blockade conductance peaks.

Zero-bias conductance can be qualitatively understood in a simple zero-temperature model in which the energy of the superconducting island—with or without sub-gap states (Fig. 1d)—is given by a series of shifted parabolas: $E_N(N_G) = E_C(N_G - N)^2 + p_N E_0$, in which $N_G = CV_G/e$ is the gate-induced charge (with electron charge e and gate capacitance C)^{19,20,22–25} and N is the electron occupancy. E_0 is the energy of the lowest quasiparticle state, which is filled for odd parity ($p_N = 1$, odd N) and empty for even parity ($p_N = 0$, even N)²¹. Transport occurs when the ground state has a charge degeneracy, that is, when the E_N parabolas intersect. For $E_0 > E_C$, the ground state always has even parity; transport in this regime occurs via tunnelling of Cooper pairs at degeneracies of the even- N parabolas. This is the regime in which the $2e$ -periodic Coulomb-blockade peaks are seen at low magnetic fields (Fig. 1d, blue). The odd charge state carries spin and its energy can be lowered by the Zeeman effect when a magnetic field is applied. For sufficiently large field, such that $E_0 < E_C$, an odd- N ground state emerges. This transition from $2e$ charging to $1e$ charging is seen experimentally as the splitting of the $2e$ -periodic Coulomb diamonds into the even-odd double-diamond pattern in Fig. 1d (green). In this regime, the Coulomb-peak spacing is proportional to $E_C + 2E_0$ for even diamonds and $E_C - 2E_0$ for odd diamonds^{23,24}. For the particular case of a zero-energy Majorana state ($E_0 = 0$) peak spacing is regular and $1e$ -periodic. This regime

¹Center for Quantum Devices and Station Q Copenhagen, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark. ²Department of Physics, Harvard University, Cambridge, Massachusetts, USA.

*These authors contributed equally to this work.

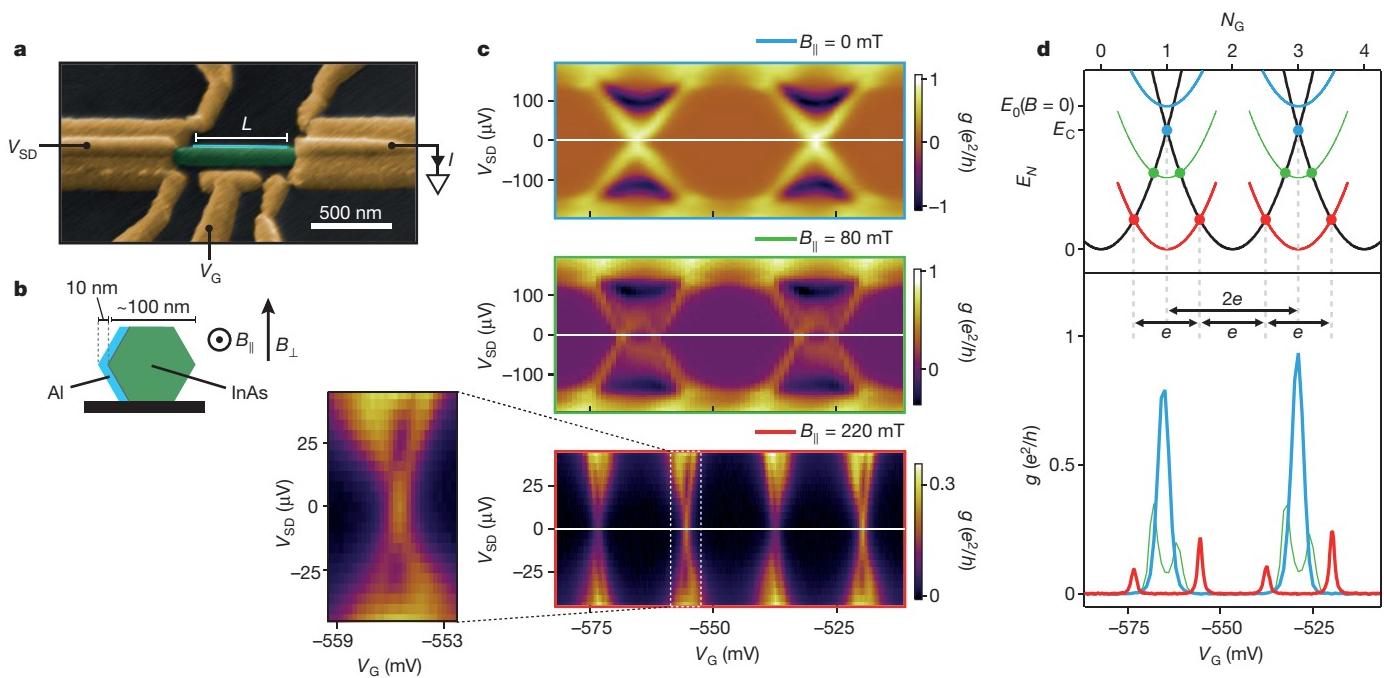


Figure 1 | Majorana island device. **a**, Electron micrograph (false colour) of a device that is lithographically similar to the measured devices. Yellow, Ti/Au contacts; green, InAs nanowire; light blue, two-facet Al shell (of length L); V_{SD} , applied voltage bias; I , measured current; V_G , gate voltage. **b**, Cross-section of a hexagonal InAs nanowire showing the orientation of the Al shell and field directions B_{\parallel} and B_{\perp} . **c**, Differential conductance $g = dI/dV_{SD}$ as a function of gate voltage V_G and source-drain bias V_{SD} for parallel magnetic fields $B_{\parallel} = \{0, 80, 220\}$ mT, showing a series of Coulomb diamonds. For $B_{\parallel} = 0$ mT, the Coulomb diamonds are evenly spaced. An odd diamond has appeared for $B_{\parallel} = 80$ mT. For $B_{\parallel} = 220$ mT, the Coulomb diamonds feature evenly spaced discrete states, but the period in gate voltage has halved compared to the $B_{\parallel} = 0$ mT case. Horizontal white lines indicate the locations of the plots shown in **d**. **d**, Top, energy E_N of the

device with electron occupancy N as a function of normalized gate voltage N_G . Ground-state energies for even (odd) N are shown in black (colour). Odd- N energies are raised by the single-particle-state energy E_0 relative to even- N energies. In regimes with even N only, $E_0 > E_C$ (in which E_C is the charging energy; light blue); in regimes with both even and odd N , $E_0 < E_C$ (green). The Majorana case ($E_0 = 0$) is shown in red. Transport can occur at the intersections of the parabolas, indicated by the filled circles. Bottom, differential conductance g versus gate voltage V_G at zero bias from measurements in **c** for magnetic fields $B_{\parallel} = \{0, 80, 220\}$ mT. The splitting of the $2e$ -periodic peak (light blue line) reflects a transition from Cooper pair tunnelling to single-quasiparticle charging of the Coulomb island. Evenly spaced, $1e$ -periodic Coulomb peaks are characteristic of a zero-energy state.

is observed at higher fields (Fig. 1d, red), although not so high as to destroy superconductivity.

Coulomb-peak spacings are measured as a function of magnetic field, allowing the state energy, $E_0(B)$, to be extracted. An example, showing ten consecutive peaks for the $L = 0.9\mu\text{m}$ device, is shown in Fig. 2a. The peaks are $2e$ -periodic at $B = 0$, start splitting at $B \approx 95$ mT and become $1e$ -periodic at $B \approx 110$ mT, well below the spectroscopically observed closing of the superconducting gap at $B_c \approx 600$ mT (see Methods). This result indicates the presence of a state close to zero energy within the superconducting regime over a range of about 500 mT.

Separately averaging even and odd Coulomb-peak spacings ($\langle S_{e,o} \rangle$) over an ensemble of adjacent peaks reveals oscillations around the $1e$ -periodic value as a function of applied magnetic field. This finding is consistent with an oscillating state energy E_0 due to hybridized Majorana modes^{13–15}. For the $L = 0.9\mu\text{m}$ device (Fig. 2b), peak-spacing oscillations yield an energy oscillation amplitude $A = 7.0 \pm 1.5\mu\text{eV}$ that is converted from gate voltage to energy using the gate lever arm η , which is extracted independently from the slope of the Coulomb diamonds. For the $L = 1.5\mu\text{m}$ device (Fig. 2c), oscillations in the average Coulomb-peak spacing determined from 22 consecutive peaks yield a barely resolvable amplitude $A = 1.2 \pm 0.5\mu\text{eV}$.

Oscillation amplitudes for the five measured devices (see Methods for device details) are shown in Fig. 2d along with a two-parameter fit to an exponential function, $A = A_0 e^{-L/\xi}$, which yields $A_0 = 300\mu\text{eV}$ and $\xi = 260\text{ nm}$ as fit parameters. The data fit well to the predicted exponential form that characterizes the topological protection of Majorana modes^{3,4,13}.

Excited states of the Majorana island are probed using finite-bias transport spectroscopy. This technique requires a fixed gate voltage,

which is chosen such that, at zero bias, the electrochemical potential of the leads aligns with the centre of the spectroscopic gap of the Majorana island. With this choice, the conductance observed at a source-drain bias V_{SD} is due to states at energy $eV_{SD}/2$. A conductance peak at zero bias corresponds to a zero-energy state. In the case shown in Fig. 3a, b, the gate voltage is tuned using the characteristic finite-bias conductance spectra for a short InAs/Al island that was investigated previously²¹. Ground-state energies determined by finite-bias spectroscopy match those extracted from zero-bias peak spacings (see Extended Data Fig. 7).

Bias spectroscopy shows discrete zero-energy states emerging at sufficient applied field over a range of device lengths. In a short device (Fig. 3c), the discrete state moves linearly as a function of magnetic field, passing through zero and merging with a continuum at $V_{SD} \approx 100\mu\text{eV}$. This merging is expected for Majorana systems in the short-length limit, in which quenching of spin-orbit coupling results in unprotected parity crossings and state intersections at high energy¹⁴. However, rather than passing directly through zero, the first zero crossing extends for 40 mT; this behaviour is not understood. For medium-length devices, the sub-gap state bends back towards zero after zero crossings (Fig. 3d), in agreement with theoretical predictions for the emergence of Majorana behaviour with increasing system length^{14,15}. For a long device ($L = 1.5\mu\text{m}$), bias spectroscopy shows a zero-energy state separated from a continuum at higher bias (Fig. 3e). The zero-energy state is present over a field range of 120 mT, with an associated energy gap of $(30\mu\text{eV})/k_B = 0.35\text{ K}$ (in which k_B is the Boltzmann constant).

The evolution with increasing device length from unprotected parity crossings to energetically isolated oscillating states and then to a fixed

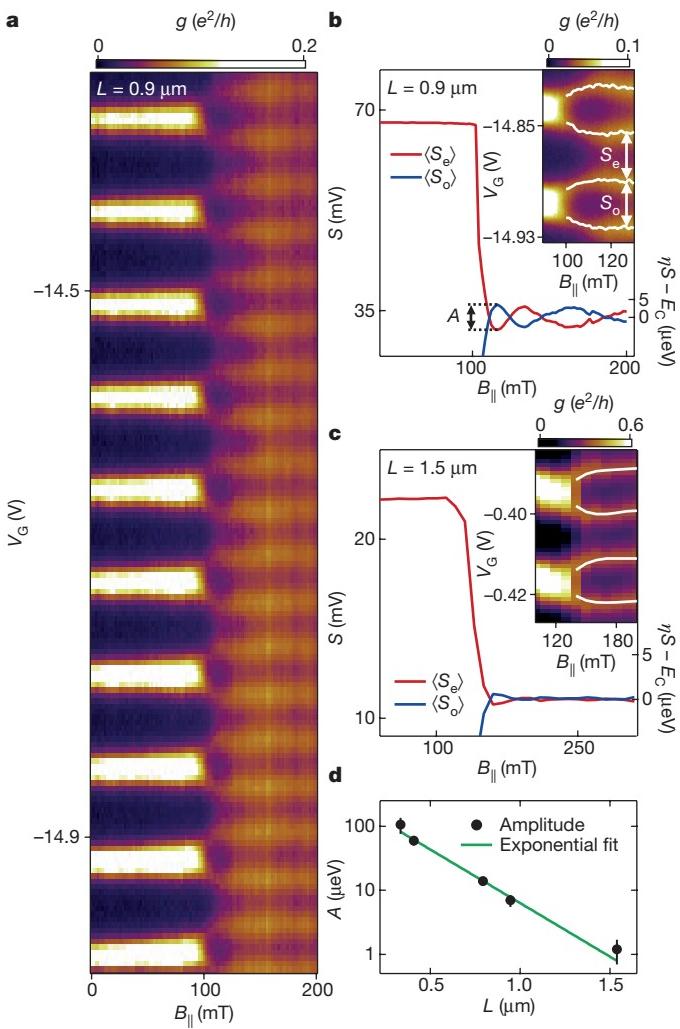


Figure 2 | Peak splitting in magnetic field. **a**, Zero-bias conductance g as a function of gate voltage V_G and parallel magnetic field B_{\parallel} for the $L = 0.9 \mu\text{m}$ device, showing a series of $2e$ -periodic Coulomb peaks below about 100 mT and nearly $1e$ -periodic peaks above about 100 mT. **b**, Average peak spacing for even and odd Coulomb valleys $\langle S_{e,o} \rangle$ from the measurement shown in **a** as a function of magnetic field B_{\parallel} . The Coulomb peaks become evenly spaced at $B_{\parallel} = 110$ mT; at higher fields, their spacing oscillates around $\langle S_e \rangle = \langle S_o \rangle$. The right axis shows the energy scale $\eta S - E_C$ in the $1e$ -periodic regime (η is gate lever arm; see text). Inset, high-resolution measurement for $L = 0.9 \mu\text{m}$ (**a**) with the peak centre overlaid. Even and odd peak spacings $S_{e,o}$ are indicated by the arrows. **c**, As for **b**, but for a longer wire, $L = 1.5 \mu\text{m}$. **d**, Oscillatory amplitude A plotted against the shell length L for 5 devices (L ranging from 330 nm to $1.5 \mu\text{m}$; black dots). The green line is an exponential fit to the data: $A = A_0 \exp(-L/\xi)$ with $A_0 = 300 \mu\text{eV}$ and $\xi = 260$ nm. Error bars indicate uncertainties propagated from lever-arm measurements and fits to peak maxima.

zero-energy state is consistent with the expected crossover from a strongly overlapping precursor of split Majorana states to a topologically protected Majorana state locked at zero energy^{14,15}. In the data in Fig. 3e, the signal from the discrete state disappears for $B_{\parallel} > 320$ mT. This is not expected for a simple (disorder-free, single sub-band) Majorana picture. Even though the zero-bias peak disappears, the peak spacing remains $1e$ -periodic (see Methods).

The observed effective g -factors of 20–50, which are extracted from the addition spectrum and bias spectroscopy (see Methods), are large compared to previous studies on InAs nanowires^{9,26,27}, perhaps as a result of field focusing from the Al shell. The measured gap to the continuum at zero magnetic field is consistent with the gap of aluminium, $\Delta_{\text{Al}} \approx 180 \mu\text{eV}$, and is roughly the same in all devices. The discrete subgap states (Fig. 3c–e) have zero-field energy that is less than,

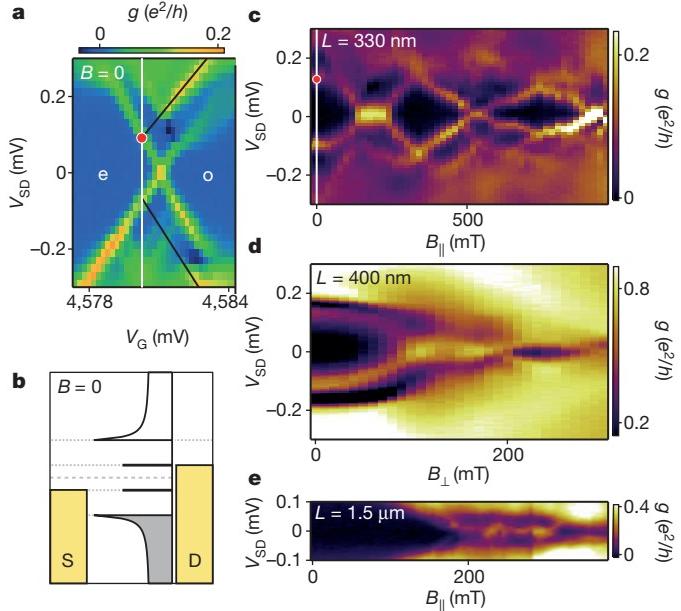


Figure 3 | Bias spectroscopy. **a**, Conductance g versus bias voltage V_{SD} and gate voltage V_G for the $L = 330 \text{ nm}$ device. Black lines indicate conductance due to a bound state; red marker is at $eV_{SD} = 2E_0$. ‘e’ and ‘o’ indicate regions in which the electron occupancy N is even and odd, respectively. **b**, Coulomb island and lead density of states at the voltage configuration indicated by the red marker in **a**. Changing the voltage bias moves the marker along the white line in **a**. ‘S’ and ‘D’ indicate the source and drain, respectively. Shading indicates occupied states. **c–e**, Conductance g versus source–drain bias V_{SD} and magnetic field B_{\parallel} (**c**, **e**) or B_{\perp} (**d**) for the $L = 330 \text{ nm}$ (**c**), $L = 400 \text{ nm}$ (**d**) and $L = 1.5 \mu\text{m}$ (**e**) devices with the gate voltage V_G equivalent to the position indicated schematically for $L = 330 \text{ nm}$ by the white line in **a**.

but comparable to, the gap, $E_0(B=0) \approx 50–160 \mu\text{eV}$, which is consistent with expectations for half-shell geometries²⁸. The measured gap between the near-zero-energy state and the continuum in the high-field (topological) regime, $\Delta_T \approx 30 \mu\text{eV}$, as well as the coherence length extracted from the exponential fit to the length-dependent splitting (Fig. 2d), $\xi \approx 260$ nm, are consistent with topological superconductivity. At low magnetic fields, the gap and coherence length are related to the strength of spin-orbit coupling: $\alpha_{SO} \approx \xi \times \Delta_T = 8 \times 10^{-2} \text{ eV } \text{\AA}$; this value is consistent with those previously reported for InAs nanowires^{9,29}. For a single sub-band, this implies a Fermi velocity $v_F = \alpha_{SO}/\hbar = 1 \times 10^4 \text{ m s}^{-1}$ that is lower than expected, suggesting that more than one sub-band is occupied under the Al shell; however, we are not able to extract the number of modes directly.

Finally, we consider the magnetic-field dependence of the heights of Coulomb-blockade peaks (as opposed to the spacings) (Fig. 4). We found in most devices that below the field B^* , at which $2e$ -periodic peaks split, all peaks had uniformly high amplitude. Above B^* , peak heights rapidly decreased and remained low up to a second characteristic field, B^{**} , coincident with the onset of $1e$ periodicity (that is, the field at which even–odd spacing differences vanished). Above B^{**} , peak heights recovered. In the longer wires, peaks were nearly absent between B^* and B^{**} (Fig. 4c).

We interpret these observations as follows. In the present lead–lead geometry, transport at fields above B^* involves single electrons entering one end of the wire and leaving from the other. The onset of uniform spacing with the reappearance of high peaks for fields above B^{**} indicates the emergence of a state (or states) at zero energy with strong wavefunction weight at both ends of the wire. This is consistent with teleportation of electrons from one end of the wire to the other via a Majorana mode^{16,17}, although it is not necessarily a unique signature of teleportation³⁰. Therefore, although the simultaneous brightening of peaks and their becoming uniformly spaced at B^{**} suggests a sub-gap or Majorana mode moving to the ends of the wire as it moves to zero

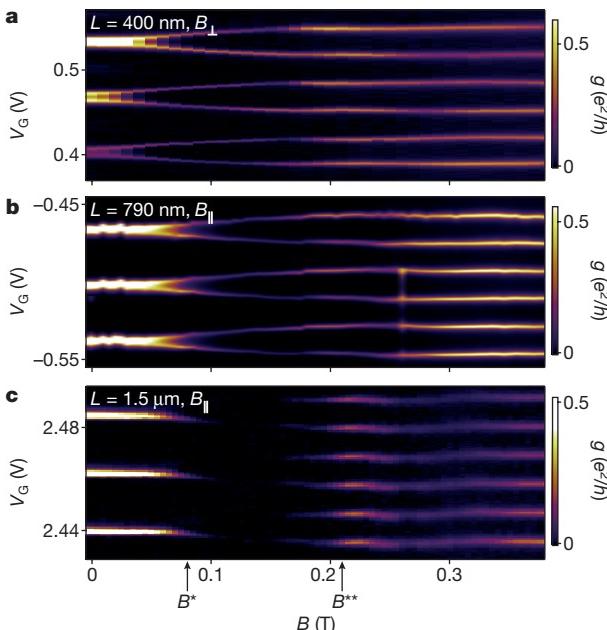


Figure 4 | Length dependence of Coulomb-peak heights. **a–c,** Conductance g as a function of magnetic field B_{\perp} (**a**) or B_{\parallel} (**b**, **c**) and gate voltage V_G for device lengths $L = 400$ nm (**a**), $L = 790$ nm (**b**) and $L = 1.5 \mu\text{m}$ (**c**). Coulomb peaks become dim at field B^* and brighten at field B^{**} , particularly for the $L = 1.5 \mu\text{m}$ device, consistent with teleportation at fields above B^{**} .

energy, we cannot rule out other forms of end-localized zero-energy states that could appear above a critical field.

In summary, we studied Majorana islands composed of InAs nanowires covered on two facets with epitaxial Al, for a range of device lengths. Zero-energy states are observed for wires of all lengths away from zero field. Oscillating energy splittings, measured using Coulomb-blockade spectroscopy, are exponentially suppressed with wire length, with a characteristic length $\xi = 260$ nm. This result constitutes an explicit demonstration of exponential protection of zero-energy modes. Finite-bias measurements show transport through a discrete zero-energy state, with a measured topological gap $\Delta_T = 30 \mu\text{eV}$ for long devices. The extracted Δ_T and ξ are consistent with known parameters for InAs nanowires and the emergence of topological superconductivity. Brightening of Coulomb peaks at the field at which spacing becomes uniform for longer devices suggests the presence of a robust delocalized state connecting the leads, and provides experimental support for electron teleportation via Majorana modes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 September 2015; accepted 18 January 2016.

- Kitaev, A. Y. Fault-tolerant quantum computation by anyons. *Ann. Phys.* **303**, 2–30 (2003).
- Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Das Sarma, S. Non-Abelian anyons and topological quantum computation. *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
- Read, N. & Green, D. Paired states of fermions in two dimensions with breaking of parity and time-reversal symmetries, and the fractional quantum Hall effect. *Phys. Rev. B* **61**, 10267–10297 (2000).
- Kitaev, A. Y. Unpaired Majorana fermions in quantum wires. *Phys. Usp.* **44**, 131–136 (2001).
- Lutchyn, R. M., Sau, J. D. & Das Sarma, S. Majorana fermions and a topological phase transition in semiconductor-superconductor heterostructures. *Phys. Rev. Lett.* **105**, 077001 (2010).
- Oreg, Y., Refael, G. & von Oppen, F. Helical liquids and Majorana bound states in quantum wires. *Phys. Rev. Lett.* **105**, 177002 (2010).

- Mourik, V. et al. Signatures of Majorana fermions in hybrid superconductor-semiconductor nanowire devices. *Science* **336**, 1003–1007 (2012).
- Rokhinson, L. P., Liu, X. & Furdyna, J. K. The fractional a.c. Josephson effect in a semiconductor-superconductor nanowire as a signature of Majorana particles. *Nature Phys.* **8**, 795–799 (2012).
- Das, A. et al. Zero-bias peaks and splitting in an Al-InAs nanowire topological superconductor as a signature of Majorana fermions. *Nature Phys.* **8**, 887–895 (2012).
- Deng, M. T. et al. Anomalous zero-bias conductance peak in a Nb-InSb nanowire-Nb hybrid device. *Nano Lett.* **12**, 6414–6419 (2012).
- Churchill, H. O. H. et al. Superconductor-nanowire devices from tunneling to the multichannel regime: zero-bias oscillations and magnetoconductance crossover. *Phys. Rev. B* **87**, 241401 (2013).
- Nadj-Perge, S. et al. Observation of Majorana fermions in ferromagnetic atomic chains on a superconductor. *Science* **346**, 602–607 (2014).
- Das Sarma, S., Sau, J. D. & Stanevici, T. D. Splitting of the zero-bias conductance peak as smoking gun evidence for the existence of the Majorana mode in a superconductor-semiconductor nanowire. *Phys. Rev. B* **86**, 220506 (2012).
- Stanevici, T. D., Lutchyn, R. M. & Das Sarma, S. Dimensional crossover in spin-orbit-coupled semiconductor nanowires with induced superconducting pairing. *Phys. Rev. B* **87**, 094518 (2013).
- Rainis, D., Trifunovic, L., Klinovaja, J. & Loss, D. Towards a realistic transport modeling in a superconducting nanowire with Majorana fermions. *Phys. Rev. B* **87**, 024515 (2013).
- Fu, L. Electron teleportation via Majorana bound states in a mesoscopic superconductor. *Phys. Rev. Lett.* **104**, 056402 (2010).
- Hützen, R., Zazunov, A., Braunecker, B., Yeyati, A. L. & Egger, R. Majorana single-charge transistor. *Phys. Rev. Lett.* **109**, 166403 (2012).
- Krogstrup, P. et al. Epitaxy of semiconductor-superconductor nanowires. *Nature Mater.* **14**, 400–406 (2015).
- Hekking, F., Glazman, L., Matveev, K. & Shekhter, R. Coulomb blockade of two-electron tunneling. *Phys. Rev. Lett.* **70**, 4138–4141 (1993).
- Hergenrother, J., Tuominen, M. & Tinkham, M. Charge transport by Andreev reflection through a mesoscopic superconducting island. *Phys. Rev. Lett.* **72**, 1742–1745 (1994).
- Higginbotham, A. P. et al. Parity lifetime of bound states in a proximitized semiconductor nanowire. *Nature Phys.* **11**, 1017–1021 (2015).
- Eiles, T. M., Martinis, J. M. & Devoret, M. H. Even-odd asymmetry of a superconductor revealed by the Coulomb blockade of Andreev reflection. *Phys. Rev. Lett.* **70**, 1862–1865 (1993).
- Tuominen, M. T., Hergenrother, J. M., Tighe, T. S. & Tinkham, M. Experimental evidence for parity-based 2e periodicity in a superconducting single-electron tunneling transistor. *Phys. Rev. Lett.* **69**, 1997–2000 (1992).
- Lafarge, P., Joyez, P., Esteve, D., Urbina, C. & Devoret, M. H. Measurement of the even-odd free-energy difference of an isolated superconductor. *Phys. Rev. Lett.* **70**, 994–997 (1993).
- Matveev, K. A., Glazman, L. I. & Shekhter, R. I. Effects of charge parity in tunneling through a superconducting grain. *Mod. Phys. Lett. B* **08**, 1007–1026 (1994).
- Csonka, S., Hofstetter, L., Freitag, F. & Oberholzer, S. Giant fluctuations and gate control of the g-factor in InAs nanowire quantum dots. *Nano Lett.* **8**, 3932–3935 (2008).
- Schroer, M. D., Petersson, K. D., Jung, M. & Petta, J. R. Field tuning the g factor in InAs nanowire double quantum dots. *Phys. Rev. Lett.* **107**, 176811 (2011).
- Cole, W. S., Das Sarma, S. & Stanevici, T. D. Effects of large induced superconducting gap on semiconductor Majorana nanowires. *Phys. Rev. B* **92**, 174511 (2015).
- Fasth, C., Fuhrer, A., Samuelson, L., Golovach, V. N. & Loss, D. Direct measurement of the spin-orbit interaction in a two-electron InAs nanowire quantum dot. *Phys. Rev. Lett.* **98**, 266801 (2007).
- Sau, J. D., Swingle, B. & Tewari, S. Proposal to probe quantum nonlocality of Majorana fermions in tunneling experiments. *Phys. Rev. B* **92**, 020511 (2015).

Acknowledgements We thank K. Flensberg, M. Leijnse, M. Deng, W. Chang and R. Lutchyn for discussions, and G. Ungaretti, S. Upadhyay, C. Sørensen, M. von Soosten and D. Sherman for contributions to growth and fabrication. This research was supported by Microsoft Project Q, the Danish National Research Foundation, the Lundbeck Foundation, the Carlsberg Foundation and the European Commission. C.M.M. acknowledges support from the Villum Foundation.

Author Contributions P.K., T.S.J. and J.N. developed the nanowire materials. S.M.A. fabricated the devices. S.M.A., A.P.H. and M.M. carried out the measurements with input from F.K., T.S.J. and C.M.M. Data analysis was done by S.M.A., A.P.H. and M.M. All authors contributed to interpreting the data. The manuscript was written by S.M.A., A.P.H. and C.M.M. with suggestions from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M.M. (marcus@nbi.dk).

METHODS

Sample preparation. The InAs nanowires with epitaxial Al shells were grown via a two-step process by molecular beam epitaxy. First, the InAs nanowires were grown using the vapour–liquid–solid method with Au as a catalyst at 420 °C. Second, after cooling the system to –30 °C, the Al was grown on two facets of the hexagonal cross-section¹⁸. Afterwards, the nanowires were deposited on degenerately doped Si substrates with 100–500-nm-thick thermal oxides using either wet or dry deposition techniques. Wet deposition involves sonicating a growth substrate of nanowires in methanol for a few seconds, then putting several drops of the nanowire–methanol solution onto the chip surface using a pipette. Dry deposition was done by bringing a small piece of cleanroom wipe in touch with the growth substrate, then swiping it onto the chip surface. We found that although wet deposition results in a more uniform dispersion of nanowires on the chip surface, dry deposition is faster and less wasteful with nanowires. Selective removal of the Al shell was done by patterning etch windows using electron beam lithography on both sides of the nanowire, plasma cleaning the surface of the nanowire using oxygen, then etching the Al using a Transene Al Etchant D with an etching time of 10 s at 50 °C. Depending on the device, ohmic contacts to the InAs core were fabricated using either ion milling or sulphur passivation to remove surface oxides. Ion milling was done for times ranging from 85 s to 110 s using a Kaufman & Robinson KDC 40 4-CM DC Ion Source with an acceleration voltage of 120 V and an ion beam current density of 0.5 mA cm^{−2} at the chip surface. Sulphur passivation was done using a 2.1% solution of (NH₄)₂S in de-ionized water with 0.15 M dissolved elemental sulphur at 40 °C for 20 min. This was followed by the deposition of 5 nm of Ti as a sticking layer and 70–100 nm of Au for the ohmic contact. We found that ion milling resulted in more stable devices. Side and plunger gates were lithographically defined in the same fabrication step as were the ohmic contacts to increase device yield. PMMA was used as resist in all lithography steps.

Device geometries. Gate patterns of the five measured devices are shown in Extended Data Fig. 1. With the exception of the $L=0.9\text{ }\mu\text{m}$ device, all measurements involving gate dependence are tuned through resonances using the plunger gate on either the Al side or the uncoated InAs side. For the $L=0.9\text{ }\mu\text{m}$ device, the lower-left side gate is used to tune through resonances of the Coulomb island, because the central plunger gate was not bonded during the cool down.

Measurements. Transport measurements were carried out in an Oxford Triton dilution refrigerator with a base electron temperature of $T \approx 50\text{ mK}$ and a 6–1 T vector magnet. Differential conductance $g = dI/dV_{SD}$ was measured using the a.c.-lock-in technique with an excitation voltage in the range 2–6 μV .

Peak spacing data summary. The exponential curve in Fig. 2d is derived from even–odd peak-spacing measurements in the high critical field directions, B_{\parallel} and B_{\perp} , summarized in Extended Data Fig. 2. Suppression of spacing fluctuations with increased device length is clearly visible. The measured amplitude A is indicated by black arrows in the insets of Extended Data Fig. 2, and the values are recorded in Extended Data Table 1 for each device length, along with charging energies and lever arms.

For $L=330\text{ nm}$, Coulomb-peak fluctuations became uncorrelated after several peaks. To obtain a large statistical ensemble, fluctuations were averaged over five sets of Coulomb peaks taken in different device tunings. Extended Data Fig. 2a shows data from a single set of peaks; Extended Data Table 1 reports the full ensemble average.

In a transverse magnetic field applied in the low critical field direction B_{tr} , shown in Extended Data Fig. 2f–i, the oscillations are absent, with the exception of an initial overshoot for $L=0.9\text{ }\mu\text{m}$ at $B_{tr}=55\text{ mT}$ (Extended Data Fig. 2i) before the system is driven into the normal state at $B_{tr} \approx 65\text{ mT}$.

Magnetic field orientation. The direction of the nanowire on the chip was found by orienting the magnetic field from a vector magnet in the chip plane and spectroscopically measuring the anisotropy of the critical magnetic field. By comparing to the wire direction on the basis of optical and electron micrographs, we estimate an angular precision of $\pm 3^\circ$.

Critical field measurements. The observed 2e-to-1e splitting at $B_{\parallel} \approx 95\text{ mT}$ is compared to the closing of the superconducting gap at a considerably higher critical field ($B_{c,\parallel}$) in Extended Data Fig. 3. Bias spectroscopy in Extended Data Fig. 3b shows a closing of the superconducting gap at $B_{c,\parallel} \approx 600\text{ mT}$, more than 500 mT after the onset of evenly spaced 1e-periodic Coulomb peaks. The change from 2e to 1e periodicity at $B_{\parallel} \approx 100\text{ mT}$ in Extended Data Fig. 3a coincides with a reduction in the measured Coulomb gap in Extended Data Fig. 3b, reflecting the transition from Cooper-pair charging (which has an energy penalty of $2E_C$) to single-electron charging (which has an energy penalty of E_C). The measurement in Extended Data Fig. 3b was taken in a Coulomb valley at the gate voltage $V_G = -14.92\text{ V}$.

Averaging of peak spacings. In Fig. 2b we show the extracted average peak spacing for several even and odd Coulomb valleys. A high-resolution measurement of the 2e-to-1e splitting is shown in Extended Data Fig. 4a. The spacings of individual even and odd valleys ($S_{e,o}$ in Extended Data Fig. 4b) exhibit the same oscillating

behaviour as the averages ($\langle S_{e,o} \rangle$), but show a small deviation from them between 100 mT and 125 mT, which might be attributable to g-factor fluctuations for successive charge occupations of the Coulomb island. Below 100 mT, the fluctuations are very small, giving an indication of instrumental noise in the measurement.

Angle dependence. The angle dependence of the anti-crossing of the state with the continuum for $L=400\text{ nm}$ is shown in Extended Data Fig. 5. We focus on magnetic fields B_{α} with angles α in the plane perpendicular to the nanowire direction. The measurements show a pronounced anti-crossing between the sub-gap state and an excitation continuum ($\alpha=112.5^\circ$ and $\alpha=135^\circ$) that is substantially reduced for $\alpha=67.5^\circ$. Interpreting angle dependence is complicated by the anisotropy of the g-factor and the critical field. The critical field is maximized for $\alpha=120^\circ$, and is reduced drastically for near-perpendicular field alignment ($\alpha=22.5^\circ$). The observed g-factors are highly dependent on field orientation and device tuning. For the $L=400\text{ nm}$ device shown in Extended Data Fig. 5, we found an approximately sinusoidal variation in g-factor by a factor of 2, with maximum g-factor occurring near $\alpha=90^\circ$.

Choice of gate voltage for bias spectroscopy. For bias spectroscopy, the gate voltage is fixed either by interpreting Coulomb diamonds, as discussed in the main text, or from even–odd peak spacings. Although details of the bias spectroscopy, such as locations of zero-crossing, depend on the choice of gate voltage, general features such as slopes, typical fluctuation amplitude and the presence of a robust excitation gap are not strongly affected by the choice of gate voltage (Extended Data Fig. 6).

Comparison of addition energies and finite-bias spectroscopy. Peak spacings are used to measure the energy of the lowest-lying state. The same information is present in the bias spectroscopy, and gives consistent results, as shown in Extended Data Fig. 7.

Bias spectroscopy of the long device. Common-mode fluctuations in Coulomb-peak position were observed in the longest ($L=1.5\text{ }\mu\text{m}$) device, as shown in Extended Data Fig. 8a. The fluctuations evidently correspond to a shift in the electrochemical potential of the dot, probably due to a nearby, field-dependent charge trap. The fluctuations are small compared to charging energy, but complicate the application of bias spectroscopy, which needs to be performed at fixed electrochemical potential. To correct for the fluctuations, we introduce an effective gate voltage

$$V_{G,\text{eff}}(B) = V_G + \delta V(B)$$

that removes the common-mode peak motion. The offset voltage is zero at low field, when Coulomb peaks are 2e-periodic ($\delta V(B)=0$ for $B \leq 175\text{ mT}$). At high field, $\delta V(B)$ is chosen so that the reference Coulomb peak (labelled in Extended Data Fig. 8b) occurs at constant $V_{G,\text{eff}}$. All non-zero $\delta V(B)$ are listed in Extended Data Table 2.

As shown in Extended Data Fig. 8b, this procedure removes the common-mode peak motion. In the case of the $L=1.5\text{ }\mu\text{m}$ device, bias spectroscopy is performed at fixed $V_{G,\text{eff}}$, which allows us to infer the energy of the sub-gap state at fixed electrochemical potential.

Zero-energy state at successive Coulomb peaks. The zero-energy state is robust over many successive Coulomb peaks, as shown in Extended Data Fig. 9. The full bias spectroscopy as a function of field is also reproducible over several peaks, as shown in Extended Data Fig. 10.

Measured g-factors. As can be seen in Extended Data Fig. 11, the state energy does not linearly depend on magnetic field. A nonlinear behaviour with magnetic field is expected in the presence of strong spin-orbit coupling and a finite critical field.

If the behaviour was strictly linear, then we would expect

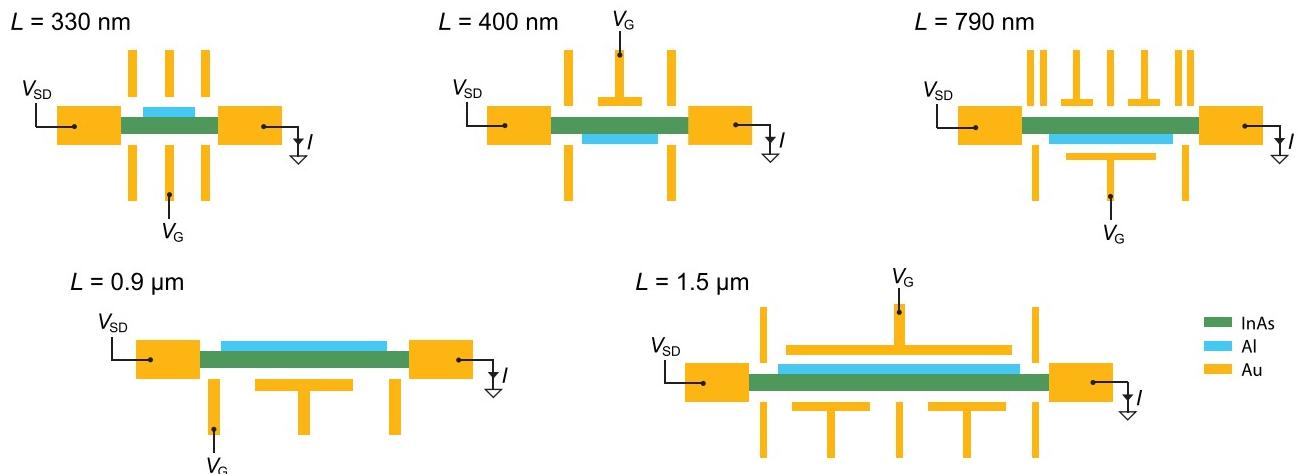
$$B^{**} = \frac{E_0}{E_0 - E_C} B^*$$

because the peak splitting at B^* occurs when $E_0(B=0) - E_Z = E_C$, and because the state is at zero energy at B^{**} when $E_Z = E_0(B=0)$, where E_Z is the Zeeman energy (see Fig. 4 for reference). The nonlinear behaviour of $E_0(B)$ at higher magnetic fields approaching B^{**} renders this approximation unsuitable for an accurate measurement of the state energy at zero field.

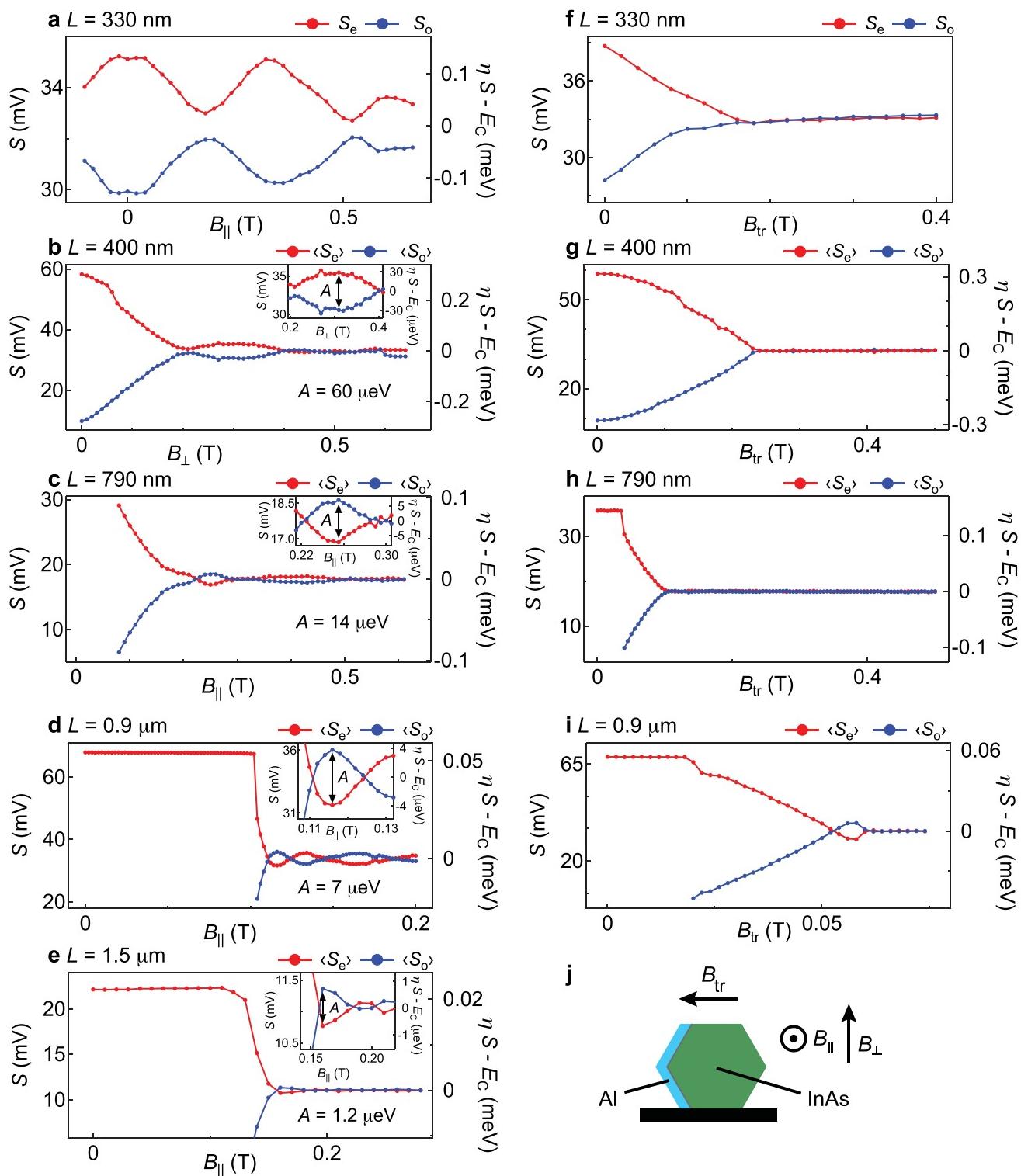
In the low-field regime in which the state energy varies approximately linearly with magnetic field, we calculate an effective g-factor. Using this slope it is possible to obtain a rough estimate of the state energy $E_0(B=0)$ assuming linear behaviour and extrapolating the state energy to zero magnetic field.

For bias spectroscopy, it should be noted that for gate voltages in the middle of the spectroscopic gap (see main text), transport through a state at $V_{SD}=V_0$ indicates a state energy $E_0 = eV_0/2$. An example for $L=330\text{ nm}$ is shown in Extended Data Fig. 11a.

Using the addition spectrum, the state energy can be calculated from the peak spacing S using $E_0 = (\eta S - E_C)/2$. Examples of extracted effective g-factors in the linear range are shown in Extended Data Fig. 11b, c.

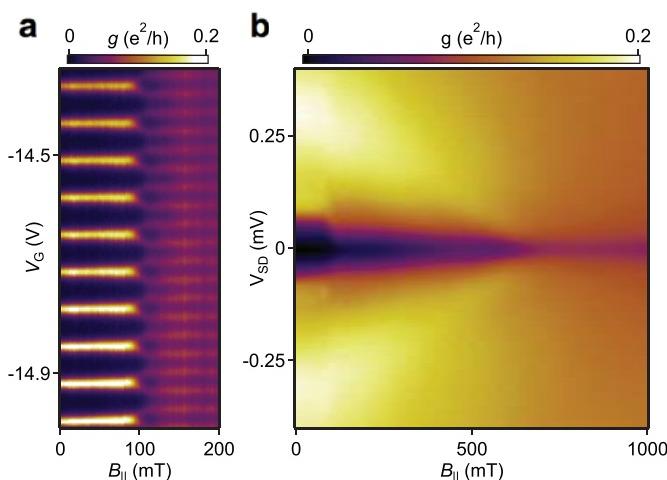


Extended Data Figure 1 | Device layouts. Gate pattern for the five measured devices showing applied voltage bias V_{SD} , measured current I and gate voltage V_G .

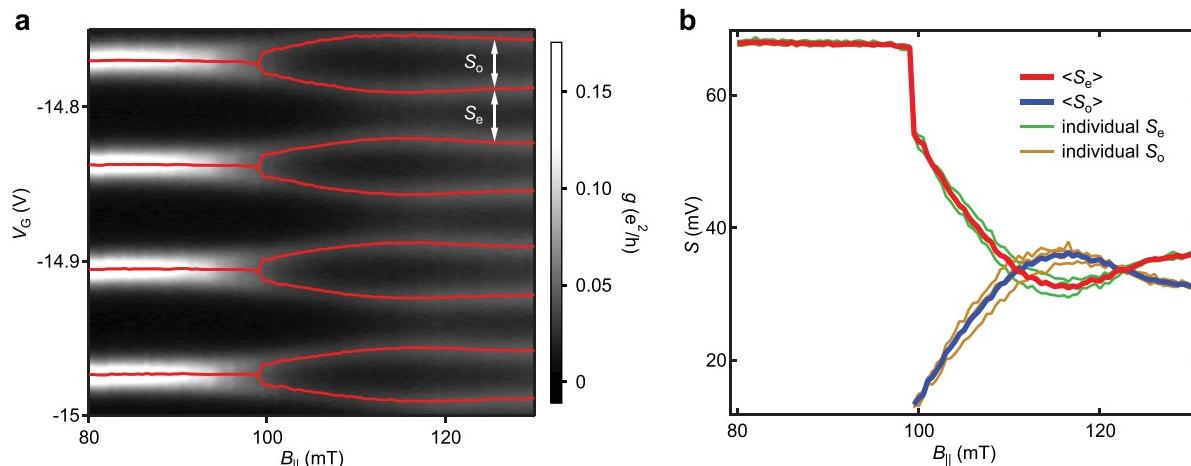


Extended Data Figure 2 | Summary of even-odd peak spacing. **a–i**, Peak spacings for even and odd valleys $S_{e,o}$ versus applied magnetic field B_{\parallel} , B_{\perp} or B_{tr} (similar to Fig. 2b) for different device lengths. Left axis shows peak spacings; right axis shows corresponding energy scales, converting from

gate voltage to energy by the lever arm η , which is measured independently from Coulomb-blockade diamonds. Insets show a magnification of the first energy splitting with an arrow indicating where A is measured. **j**, Cross-section of the nanowire, showing the applied field directions B_{\parallel} , B_{\perp} and B_{tr} .

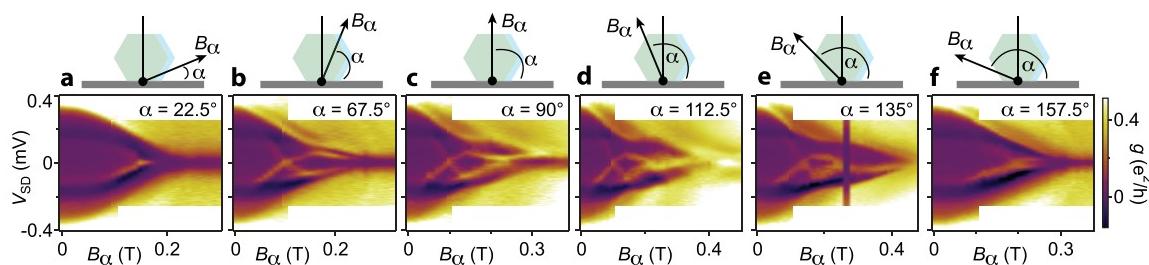


Extended Data Figure 3 | Critical field measurement for the $L = 0.9 \mu\text{m}$ device. **a**, Conductance g versus gate voltage V_G and parallel magnetic field B_{\parallel} at zero bias showing the $2e$ -to- $1e$ peak splitting. **b**, Conductance versus source-drain voltage V_{SD} and B_{\parallel} , taken at $V_G = -14.92$ V, showing a closing of the superconducting gap at $B_c \approx 640$ mT, more than 500 mT after the onset of $1e$ periodicity.

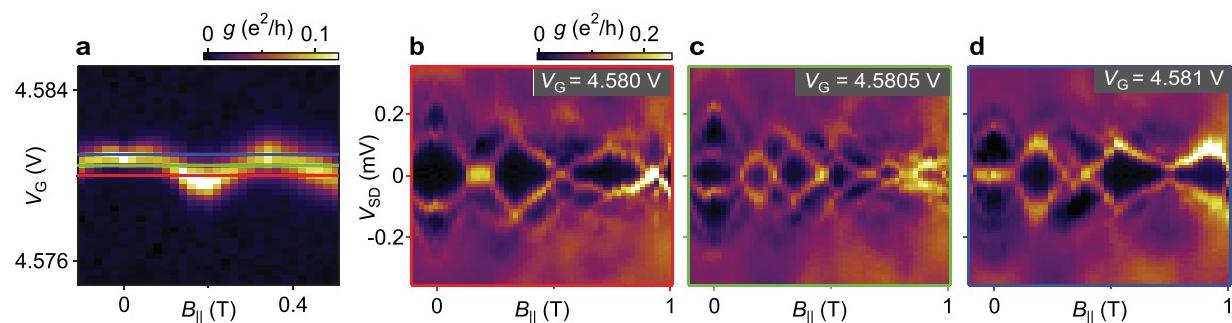


Extended Data Figure 4 | Oscillating 1e-periodic peak spacings. **a**, Zero-bias conductance g versus gate voltage V_G and parallel magnetic field B_{\parallel} at zero bias showing the 2e-to-1e peak splitting for $L = 0.9 \mu\text{m}$. The fitted peak position is indicated by a red line; even and odd peak spacings $S_{e,o}$

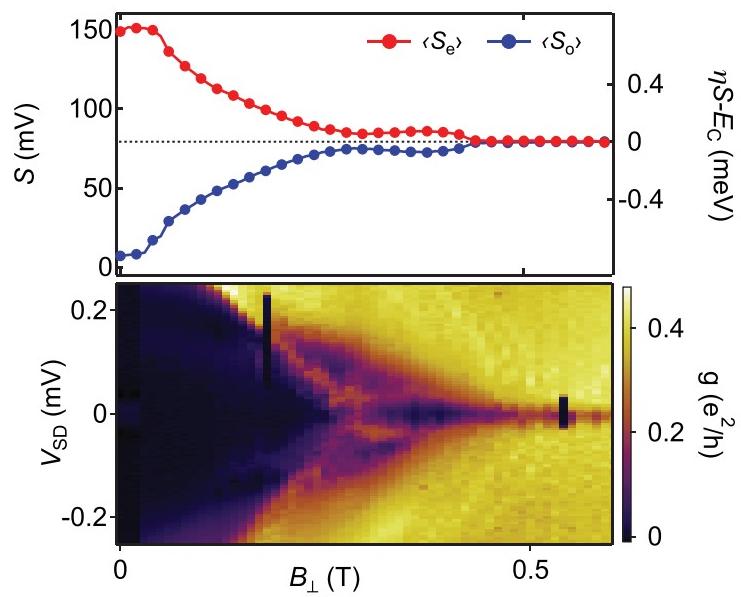
are indicated by white arrows. **b**, Peak spacing for even and odd valleys as a function of B_{\parallel} . The plot shows the average peak spacings $\langle S_{e,o} \rangle$ as well as the individual peak spacings $S_{e,o}$.



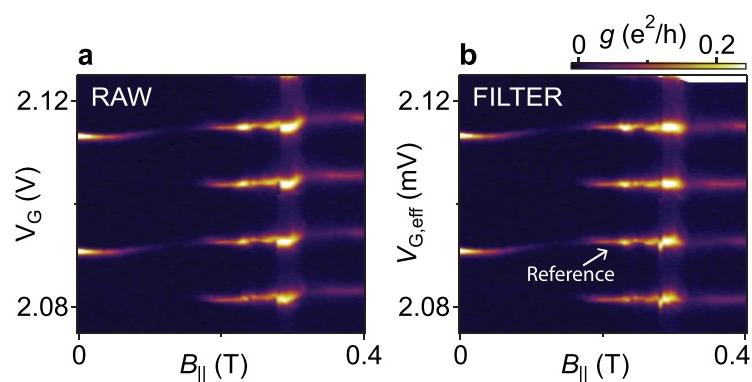
Extended Data Figure 5 | Angle dependence of state-continuum anti-crossing. **a–f**, Differential conductance g as a function of source–drain bias V_{SD} and magnetic field B_α for different angles ($\alpha = 22.5^\circ$ – 157.5°) in the plane perpendicular to the nanowire direction. Measurements are from the $L = 400$ nm device.



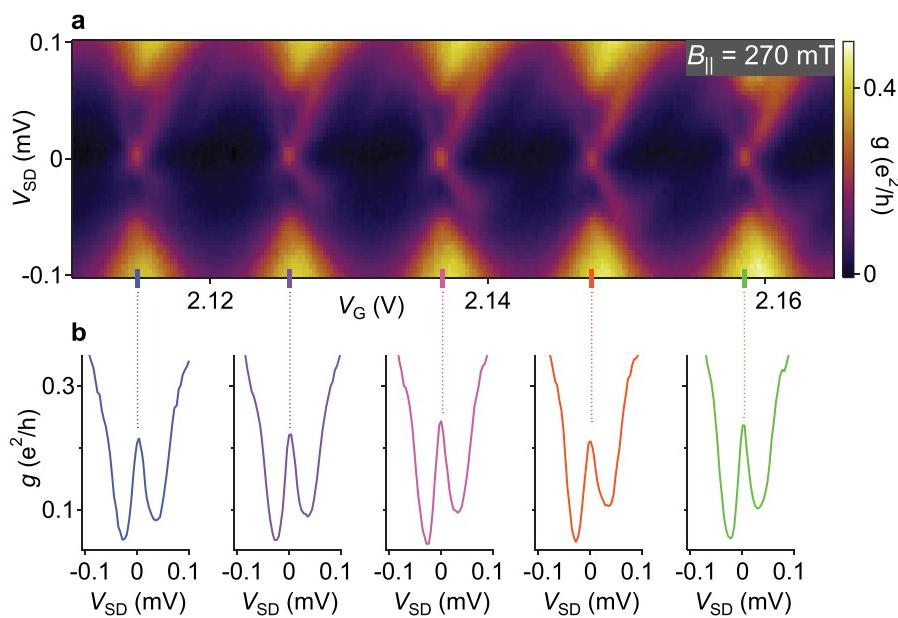
Extended Data Figure 6 | Gate positions. **a**, Differential conductance g as a function of gate voltage V_G and parallel magnetic field B_{\parallel} for the $L = 330$ nm device. Three different gate positions are indicated by coloured horizontal lines. **b–d**, Differential conductance as function of bias voltage V_{SD} and B_{\parallel} for the three gate voltages in **a**.



Extended Data Figure 7 | Comparison of peak spacings and bias spectroscopy. **a**, Peak spacing for even and odd valleys $\langle S_{e,o} \rangle$ versus applied field B_\perp . **b**, Differential conductance g as a function of source–drain bias V_{SD} and magnetic field B_\perp .

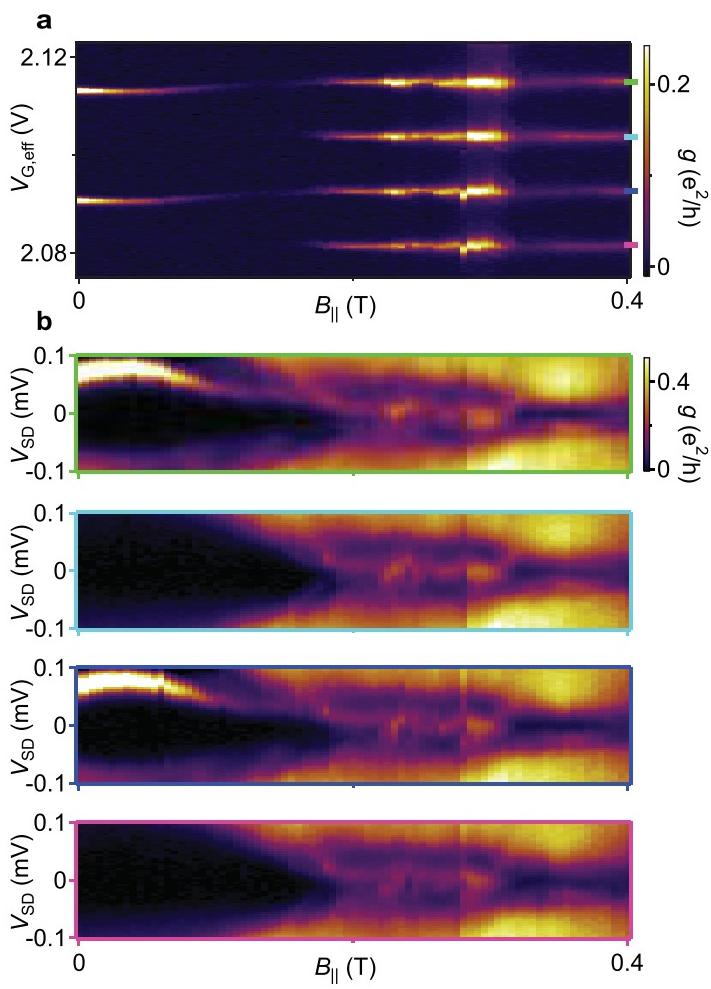


Extended Data Figure 8 | Common-mode peak motion removal. **a**, Differential conductance g versus gate voltage V_G and applied magnetic field B_{\parallel} for the $L = 1.5 \mu\text{m}$ device. **b**, Same as **a**, but with effective gate voltage $V_{G,\text{eff}}$ defined to remove common-mode peak motion. The reference Coulomb peak that is used for common-mode removal is labelled.



Extended Data Figure 9 | Zero-energy state. **a**, Differential conductance g as a function of bias voltage V_{SD} and gate voltage V_{G} for the $L = 1.5 \mu\text{m}$ device and $B_{\parallel} = 270 \text{ mT}$, showing an evenly spaced Coulomb diamond pattern and the associated gapped zero-energy state. **b**, Differential

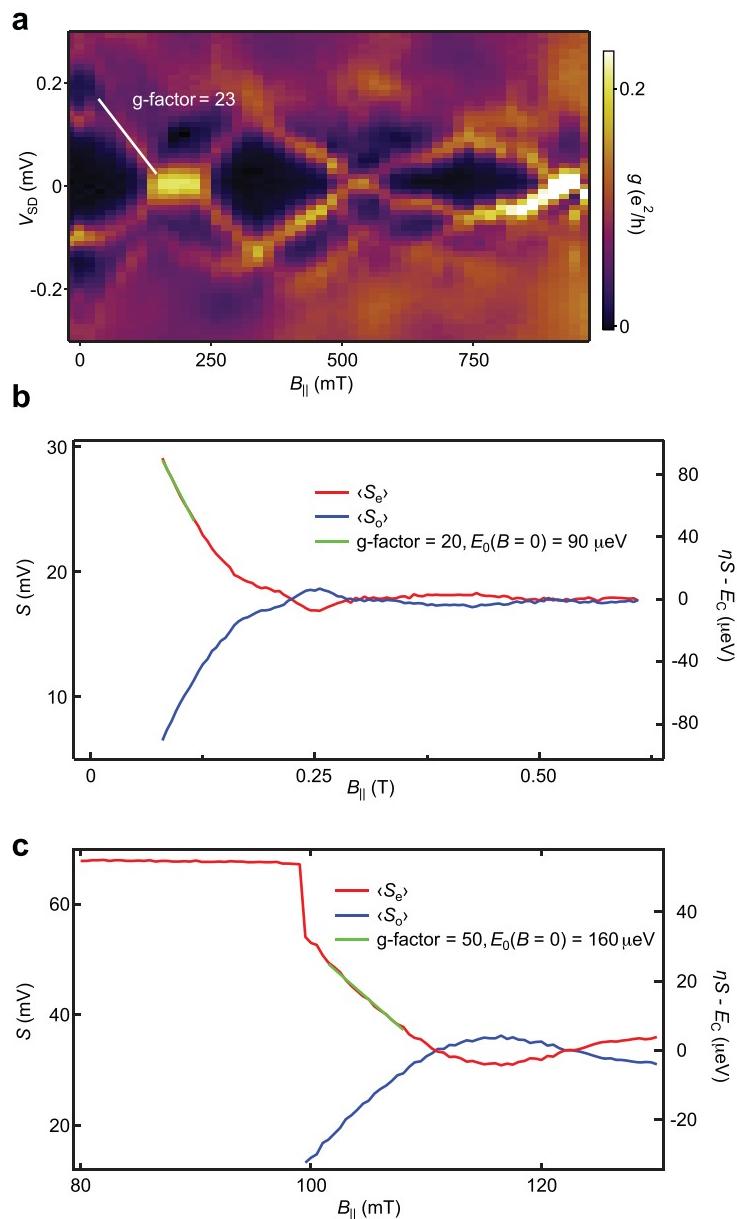
conductance versus bias voltage at the gate voltages indicated by coloured ticks in **a**. At these V_{G} values, the presence of a zero-energy state is indicated by a zero-bias peak.



Extended Data Figure 10 | Bias-spectroscopy at successive Coulomb peaks. **a**, Differential conductance g versus effective gate voltage $V_{G,\text{eff}}$ and applied magnetic field B_{\parallel} . $V_{G,\text{eff}}$ is defined to remove common-mode peak

motion; see Methods section ‘Bias spectroscopy of the long device’.

b, Differential conductance versus source-drain bias V_{SD} and applied magnetic field B_{\parallel} at fixed $V_{G,\text{eff}}$ indicated by the coloured ticks on the right axis of **a**.



Extended Data Figure 11 | Measurement of the g-factor for three devices. **a**, Differential conductance g versus source–drain voltage V_{SD} and applied magnetic field B_{\parallel} for the $L = 330$ nm device, showing a g-factor of 23.

b, c, Average even and odd peak spacings $\langle S_{e,o} \rangle$ as a function of B_{\parallel} for the $L = 790$ nm and $L = 0.9 \mu\text{m}$ devices, showing extracted g-factors of 20 and 50, respectively.

Extended Data Table 1 | Device length L , charging energy E_C , lever arm η and characteristic amplitude A for the five measured devices

L [nm]	E_C [meV]	η [eV/V]	A [μ eV]
330	1.6	0.048	106
400	0.40	0.012	60
790	0.14	0.008	14
950	0.054	0.0016	7
1540	0.022	0.002	1.2

Extended Data Table 2 | All non-zero offset voltage values $\delta V(B)$ for the $L=1.5\mu\text{m}$ device

B (mT)	δV (mV)
180	0.25
230	0.25
235	0.25
240	0.25
245	0.25
250	0.25
255	0.25
260	0.5
265	0.5
270	0.5
275	0.5
280	0.75
300	0.25
305	0.75
310	1.25
315	1.5
320	1.75
325	1.75
330	1.75
335	1.75
340	1.75
345	1.75
350	1.75
355	1.75
360	1.75
365	1.75
370	1.75
375	1.75
380	1.75
385	1.75
390	1.75
395	1.75
400	1.75

Offset is defined for $B=\{0, 5, 10, \dots, 400\}$ mT.

Change of carrier density at the pseudogap critical point of a cuprate superconductor

S. Badoux¹, W. Tabis^{2,3}, F. Laliberte², G. Grissonnanche¹, B. Vignolle², D. Vignolles², J. Béard², D. A. Bonn^{4,5}, W. N. Hardy^{4,5}, R. Liang^{4,5}, N. Doiron-Leyraud¹, Louis Taillefer^{1,5} & Cyril Proust^{2,5}

The pseudogap is a partial gap in the electronic density of states that opens in the normal (non-superconducting) state of cuprate superconductors and whose origin is a long-standing puzzle. Its connection to the Mott insulator phase at low doping (hole concentration, p) remains ambiguous¹ and its relation to the charge order^{2–4} that reconstructs the Fermi surface^{5,6} at intermediate doping is still unclear^{7–10}. Here we use measurements of the Hall coefficient in magnetic fields up to 88 tesla to show that Fermi-surface reconstruction by charge order in the cuprate $\text{YBa}_2\text{Cu}_3\text{O}_y$ ends sharply at a critical doping $p = 0.16$ that is distinctly lower than the pseudogap critical point $p^* = 0.19$ (ref. 11). This shows that the pseudogap and charge order are separate phenomena. We find that the change in carrier density n from $n = 1 + p$ in the conventional metal at high doping (ref. 12) to $n = p$ at low doping (ref. 13) starts at the pseudogap critical point. This shows that the pseudogap and the antiferromagnetic Mott insulator are linked.

Electrons in cuprate materials go from a correlated metallic state at high p to a Mott insulator at $p = 0$. How the system evolves from one state to the other remains a fundamental question. At high doping, the Fermi surface of cuprates is well established. It is a large hole-like cylinder whose volume yields a carrier density $n = 1 + p$, as measured, for example, by quantum oscillations¹⁴, in agreement with band structure calculations. The carrier density can also be measured using the Hall coefficient R_H , because in the limit of $T = 0$ the Hall number n_H of a single-band metal is such that $n_H = n$. Indeed, in the cuprate $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ (Tl-2201), the normal-state Hall coefficient R_H at $p \approx 0.3$, measured at $T \rightarrow 0$ in magnetic fields large enough to suppress superconductivity, is such that $n_H = V/(eR_H) \approx 1 + p$, where e is the electron charge and V the volume per Cu atom in the CuO_2 planes^{12,15}.

By contrast, at low doping, measurements of R_H in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (LSCO) (ref. 13) and $\text{YBa}_2\text{Cu}_3\text{O}_y$ (YBCO) (ref. 16) yield $n_H \approx p$, below $p \approx 0.08$. Having a carrier density equal to the hole concentration, $n = p$, is known to be an experimental signature of the lightly doped cuprates. The question is: at what doping does the transition between those two limiting regimes take place? Specifically, does the transition from $n = 1 + p$ to $n = p$ occur at p^* , the critical doping for the onset of the pseudogap phase? The pseudogap is a partial gap in the normal-state density of states that appears below $p^* \approx 0.19$ (ref. 11), and whose origin is a central puzzle in the physics of correlated electrons and the subject of much debate.

To answer this question using Hall measurements, one needs to reach low temperatures, which requires the use of large magnetic fields to suppress superconductivity. The only prior high-field study of cuprates that goes across p^* was performed on LSCO (ref. 17), a cuprate superconductor with a relatively low critical temperature ($T_c < 40$ K) and critical field ($H_c < 60$ T). For mainly two reasons, studies on LSCO were inconclusive on the transition from $n = 1 + p$ to $n = p$. First, the

Fermi surface of overdoped LSCO undergoes a Lifshitz transition from a hole-like to an electron-like surface as its band structure crosses a saddle-point van Hove singularity at $p \approx 0.2$ (ref. 18). This transition causes large changes in $R_H(T)$ (ref. 15) that can mask the effect of the pseudogap onset at $p^* \approx 0.19$. The second reason is the ill-defined impact of the charge-density-wave (CDW) modulations that develop at low temperature in a doping range near $p \approx 0.12$ (ref. 19). Such CDW modulations should cause a reconstruction of the Fermi surface, and hence change R_H at low temperature⁶. Therefore, the anomalies in n_H versus p observed below 60 K in LSCO (ref. 17)—and in $\text{Bi}_2\text{La}_{2-x}\text{Sr}_x\text{CuO}_{6+\delta}$ (ref. 20)—between $p \approx 0.1$ and $p \approx 0.2$ are most likely to be the combined result of three effects that have yet to be disentangled: Lifshitz transition, Fermi-surface reconstruction (FSR) and pseudogap.

Here we turn to YBCO, a cuprate material with several advantages. First, it is one of the cleanest and best ordered of all cuprates, thereby ensuring a homogeneous doping ideal for distinguishing nearby critical points. Second, the location of the pseudogap critical point is well established in YBCO, at $p^* = 0.19 \pm 0.01$ (ref. 11). Third, the Lifshitz transition in YBCO occurs at $p > 0.29$ (ref. 21), well above p^* . Fourth, the CDW modulations in YBCO have been thoroughly characterized. They are detected by X-ray diffraction (XRD) between $p \approx 0.08$ and $p \approx 0.16$ (refs 22, 23), below a temperature T_{XRD} (Fig. 1a). Above a threshold magnetic field, CDW order is detected by NMR (refs 2, 24) below a temperature T_{NMR} (Fig. 1b). Fifth, the FSR caused by the CDW modulations has a well-defined signature in the Hall effect of YBCO: $R_H(T)$ decreases smoothly to become negative at low temperature⁶—the signature of an electron pocket in the reconstructed Fermi surface. Prior Hall measurements in magnetic fields up to 60 T show that the CDW-induced FSR begins sharply at $p = 0.08$ and persists up to $p = 0.15$, the highest doping reached so far⁶.

YBCO has one disadvantage, however. Its orthorhombic structure contains conducting CuO chains along the b axis, which reduce the Hall signal coming from the CuO_2 planes. While this has no impact on the qualitative features of $R_H(T)$ (such as its sign or its qualitative T dependence), it does modify the quantitative relation between the measured Hall number n_H and the inferred carrier density n . Specifically, $n = (\rho_b/\rho_a)n_H$ (ref. 16), where ρ_b and ρ_a are the in-plane resistivities parallel and perpendicular to the b axis, respectively (see Methods and Extended Data Fig. 1).

We have performed Hall measurements in YBCO up to 88 T, allowing us to extend the doping range upwards, and hence track the normal-state properties across p^* , down to at least $T = 40$ K. Our complete data on four YBCO samples with dopings $p = 0.16, 0.177, 0.19$ and 0.205 are displayed in Extended Data Figs 2, 3, 4 and 5, respectively. In Fig. 2, we compare field sweeps of R_H versus H at $p = 0.15$ (Fig. 2a; from ref. 6) and $p = 0.16$ (Fig. 2b), at various temperatures down to 25 K. The difference is striking. At $p = 0.15$, the high-field isotherms $R_H(H)$ drop monotonically with decreasing T until they become negative at low T .

¹Département de physique, Regroupement Québécois sur les Matériaux de Pointe, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada. ²Laboratoire National des Champs Magnétiques Intenses (CNRS, EMFL, INSA, UGA, UPS), Toulouse 31400, France. ³AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, 30-059 Krakow, Poland. ⁴Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. ⁵Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8, Canada.

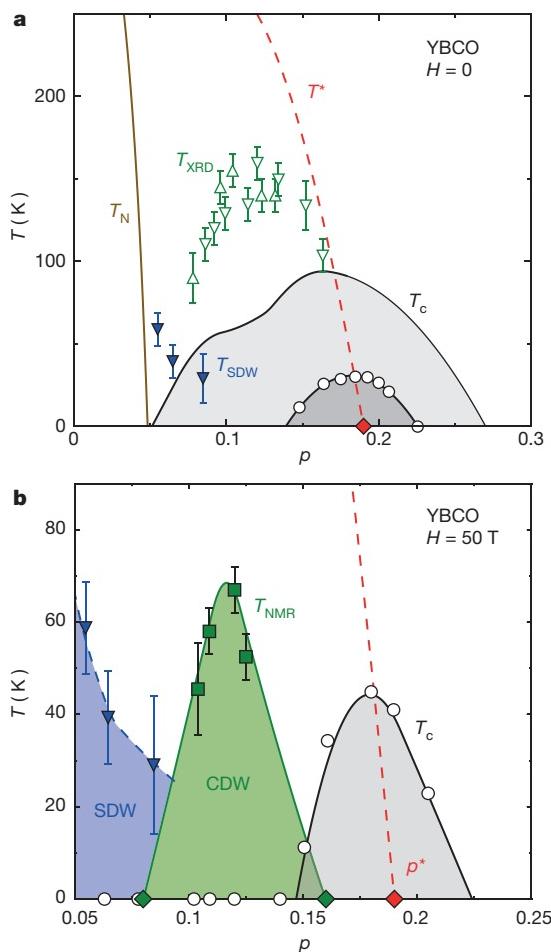


Figure 1 | Temperature-doping phase diagram of YBCO. **a**, Phase diagram in zero magnetic field ($H=0$). The superconducting phase (grey dome) lies below T_c (solid black line) and the antiferromagnetic phase lies below T_N (brown line). The small (dark grey) dome shows how T_c is suppressed by substituting 6% of the Cu atoms for Zn (white circles from ref. 25). Short-range charge-density-wave (CDW) modulations are detected by X-ray diffraction below T_{XRD} (upward-pointing open triangles and error bars from ref. 22; downward-pointing open triangles and error bars from ref. 23). Note that unlike T_{XRD} , the amplitude of the CDW modulations decreases monotonically to zero as doping goes from $p=0.12$ to $p_{\text{CDW}}=0.16 \pm 0.005$ (refs 22, 23). Short-range spin-density-wave (SDW) modulations are detected by neutron diffraction below T_{SDW} (blue triangles and error bars²⁸). The red dashed line marks the approximate location of the pseudogap temperature T^* , while $p^*=0.19 \pm 0.01$ marks the critical doping below which the pseudogap is known to appear¹¹ (red diamond). **b**, Phase diagram in a magnetic field of $H=50$ T. Above a threshold magnetic field, CDW order is detected by NMR (ref. 2) below a transition temperature T_{NMR} (green squares and error bars²⁴). The green region is where the Hall coefficient R_H is negative (from ref. 6 and this work), starting above $p=0.08$ (left green diamond). Our Hall data show that Fermi-surface reconstruction, and hence CDW order, ends at $p_{\text{FSR}}=0.16 \pm 0.005$ (right green diamond). The red dashed line is the same as in **a**. The zero-field SDW phase is reproduced from **a** (blue region). The black and green solid lines and the blue dashed line are guides to the eye.

At $p=0.16$, $R_H(H)$ never drops. Figure 3 compares the temperature evolution of the normal-state R_H at different dopings. In Fig. 3a, we see that $R_H(T)$ at $p=0.16$ shows no sign of the drop to negative values displayed at $p=0.12, 0.135$ and 0.15 , at least down to $T=40$ K. Having said this, and although the isotherms at $T=25$ K and 30 K are consistent with a constant R_H below $T=50$ K (Fig. 2), we cannot exclude that $R_H(T)$ might drop below 40 K. However, even if it did, the onset temperature for FSR would have to be much lower than it is at $p=0.15$, and it would extrapolate to zero at $p<0.165$ (Extended Data Fig. 6). We therefore

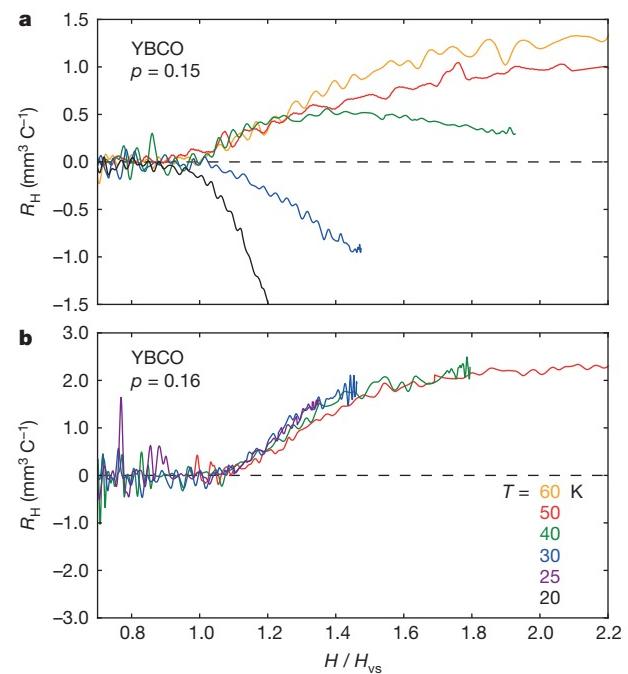


Figure 2 | Field dependence of the Hall coefficient in YBCO. **a, b**, Hall coefficient (R_H) of YBCO at various fixed temperatures, as indicated, plotted as R_H versus H/H_{vs} , where $H_{\text{vs}}(T)$ is the vortex-solid melting field above which R_H becomes non-zero, for two dopings: $p=0.15$ (**a**) and $p=0.16$ (**b**). Upon cooling, we see that R_H decreases and eventually becomes negative at $p=0.15$, while it never drops at $p=0.16$.

find that the critical doping above which there is no FSR in the normal state of YBCO at $T=0$ is $p_{\text{FSR}}=0.16 \pm 0.005$. Because this is in excellent agreement with the maximal doping at which short-range CDW modulations have been detected by XRD, namely $p_{\text{XRD}}=0.16 \pm 0.005$ (refs 22, 23), and it is consistent with the region of CDW order seen by NMR (ref. 24) (Fig. 1b), we conclude that the critical doping where CDW order ends in YBCO is $p_{\text{CDW}}=0.16 \pm 0.005$. This is consistent with the highest doping at which quantum oscillations from the CDW-induced electron pocket have been detected, namely $p=0.152$ (ref. 9).

An onset of CDW order at $p_{\text{CDW}}=0.16$ is distinctly lower than the onset of the pseudogap. Indeed, extensive analysis of the normal-state properties of YBCO above T_c yields $p^*=0.19 \pm 0.01$ (ref. 11). The critical point p^* can also be located by suppressing superconductivity with 6% Zn impurities²⁵, which shrinks the T_c dome to a small region centred around $p^*=0.19$ (Fig. 1a). This robustness of p^* confirms that CDW order and pseudogap are distinct phenomena, since CDW modulations are rapidly weakened by Zn substitution²⁶. Applying a field of 50 T produces a small T_c dome peaked at exactly the same doping, showing that $p^*=0.19 \pm 0.01$ in the normal state, whether induced by Zn or by field (Fig. 1).

We have arrived at our first main finding: the onset of pseudogap and CDW order occurs at two distinct and well-separated critical dopings. Just as $T_{\text{XRD}} < T^*$ (and $T_{\text{NMR}} < T^*$) (Fig. 1), we now find that $p_{\text{CDW}} < p^*$, in the normal state of YBCO. This contrasts with the simultaneous onset of pseudogap and short-range CDW modulations observed in the zero-field superconducting state of $\text{Bi}_2\text{Sr}_2\text{CaCuO}_{8+x}$ (Bi-2212) by scanning tunnelling microscopy (ref. 8).

Having established that the FSR due to CDW order ends at $p_{\text{FSR}}=0.16$, we now see what happens at higher p . At $p=0.205$, the temperature dependence of R_H in YBCO is similar to that of Tl-2201 (refs 12, 15) at dopings where the Fermi surface is known to be a single large hole-like cylinder with carrier density $n=1+p$ (refs 14, 15) (Extended Data Fig. 7). In particular, as T increases from zero, $R_H(T)$ rises initially, because of the growth in inelastic scattering, which is anisotropic around the large Fermi surface¹⁵. This yields a characteristic

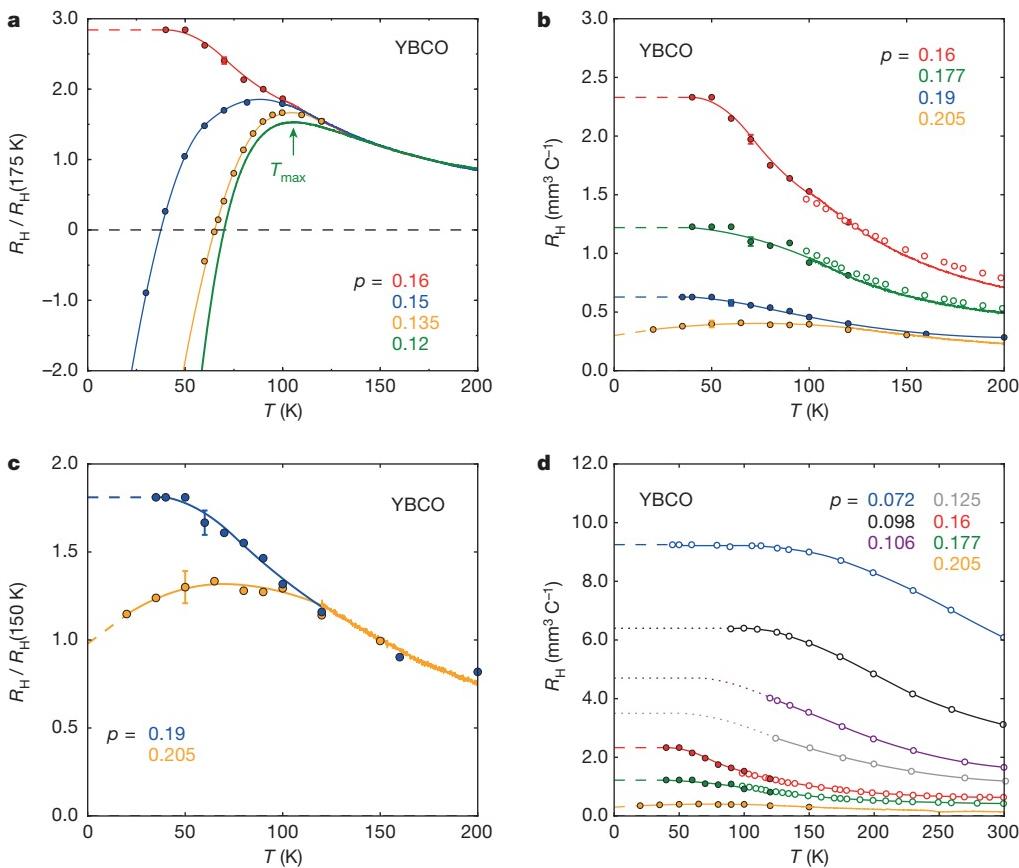


Figure 3 | Temperature dependence of the normal-state Hall coefficient in YBCO at various dopings. **a**, Data points (circles), R_H normalized by its value at $T = 175$ K. The solid curves are temperature sweeps at $H = 16$ T, above $T = 100$ K (red), 100 K (blue), 120 K (yellow) and 60 K (green). Solid lines are guides through the data points, below 100 K (red and blue) and 120 K (yellow). The red dashed line is a flat extrapolation below 40 K. The data points for $p = 0.16$ (red) are taken at (or extrapolated to) $H = 80$ T, from the R_H versus H isotherms in Extended Data Fig. 2. The data points for $p = 0.15$ (blue) and $p = 0.135$ (yellow) are taken from ref. 6 (at $H = 55$ T). The arrow marks the location of the peak in R_H versus T , for $p = 0.12$ (T_{\max}). The drop in $R_H(T)$ at low temperature is the signature of Fermi-surface reconstruction (FSR), caused by charge-density-wave (CDW) order. At $p = 0.16$, no such drop occurs, at least down to 40 K. This reveals that the critical doping for the end of FSR in the doping phase diagram (Fig. 1b) is $p_{\text{FSR}} = 0.16 \pm 0.005$ (Extended Data Fig. 6). **b**, R_H versus T at $p = 0.16$ and higher, measured at (or extrapolated to) $H = 80$ T (filled circles), from isotherms in Extended Data Figs 2, 3, 4 and 5. The curves are temperature sweeps at $H = 16$ T, above $T = 100$ K (red), 100 K (green), 120 K (yellow).

peak in $R_H(T)$, at $T \approx 100$ K (Extended Data Fig. 7). Moving to $p = 0.19$, a qualitative change has taken place (Fig. 3c): $R_H(T)$ now shows no sign of a decrease as $T \rightarrow 0$, down to our lowest temperature of 35 K (Extended Data Fig. 8). The extrapolated $T = 0$ value, $R_H(0)$, doubles upon crossing p^* .

Moving to still lower doping, we see that there is also a major quantitative change: the magnitude of R_H at low T undergoes a nearly sixfold increase between $p = 0.205$ and $p = 0.16$ (Fig. 3b), seen directly in the raw data at $T = 50$ K (Fig. 4a). We attribute this huge increase in R_H to a corresponding decrease in carrier density. In other words, states at the Fermi surface are lost and $R_H(T=0)$ increases. One may argue that for $p < 0.2$ $R_H(T)$ could decrease below 50 K and reach a value at $T = 0$ such that $n_H = 1 + p$ for all dopings down to $p = 0.16$. In this scenario, the peak in $R_H(T)$ at $T = 50$ K would be due to an anisotropic inelastic scattering that grows rapidly with underdoping¹⁵. In Methods and Extended Data Fig. 9, we show that this mechanism is inconsistent with the measured resistivity of YBCO, which is essentially independent of doping at $T = 50$ K (Extended Data Fig. 9).

Colour-coded lines are a guide to the eye through the data points. The dashed lines are a linear extrapolation below the lowest data point. Open circles are low-field data from ref. 16 for the normal-state $R_H(T)$ of YBCO above T_c , for $p = 0.16$ (red, $y = 6.95$, $T_c = 93$ K) and $p = 0.178$ (green, $y = 7.00$, $T_c = 91$ K). These data are in excellent quantitative agreement with our own data. The error bars reflect the relative uncertainty in determining the change in R_H versus T for a given doping (see Methods). Shown only for one data point per doping, the colour-coded error bar is the same for all points on the corresponding curve (doping). **c**, Same as **b** showing the two highest dopings only, but with R_H normalized at $T = 150$ K. The curve at $p = 0.19$ is qualitatively different from the curve at $p = 0.205$, showing no sign of a drop at low T (Extended Data Fig. 8). We attribute the twofold increase in the magnitude of R_H at $T \rightarrow 0$ to a decrease in carrier density as the pseudogap opens at p^* , with p^* located between $p = 0.205$ and $p = 0.19$. **d**, Same as **b** but over a wider range of doping and temperature. For the three curves in the interval $0.09 < p < 0.15$, the dotted lines show how the normal-state $R_H(T)$ might extend down to $T = 0$ in the absence of the FSR caused by CDW order.

In Fig. 4b, we plot n_H versus p and discover that in the normal state of YBCO the transition from the conventional metal at high p (where $n_H = 1 + p$) to the lightly doped regime at low p (where $n_H = p$) starts sharply at $p = p^*$, where the pseudogap opens. This is our second main finding. The observed change in R_H by a factor of ~ 6 is now understandable, since $(1 + p^*)/p^* = 6.3$. It is important to note that the huge rise in $R_H(0)$ as p is reduced below p^* is the result of a gradual process that begins at high temperature. As seen in Fig. 3d, the order-of-magnitude growth in R_H with decreasing p seen at $T \rightarrow 0$ is also observed at 300 K. Moreover, this growth is monotonic. Those two facts are consistent with the pseudogap phase, whose characteristic temperature T^* rises monotonically with decreasing p , to values exceeding 300 K (Fig. 1a). By contrast, CDW modulations cannot be responsible for the enhanced $R_H(T)$, since their onset temperature is non-monotonic and it never exceeds 150 K (Fig. 1a).

In the pseudogap phase, the topology of the $T = 0$ Fermi surface in the absence of superconductivity and CDW order is unknown. However, because the pseudogap opens at reciprocal-space locations

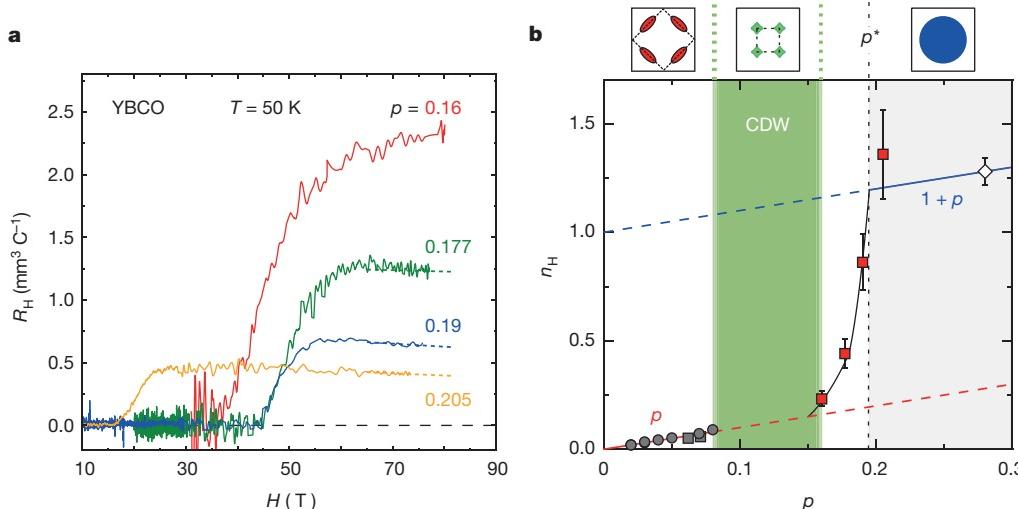


Figure 4 | Doping evolution of the normal-state carrier density.

a, Isotherms of R_H versus H in YBCO at $p = 0.16, 0.177, 0.19$ and 0.205 , measured at $T = 50$ K. Note the huge increase in the value of R_H at $H = 80$ T (or extrapolated to $H = 80$ T; dashed lines), by a factor 5.7, when going from $p = 0.205$ to $p = 0.16$. **b**, Doping dependence of the Hall number, $n_H = V/(eR_H)$, in hole-doped cuprates, measured in the normal state at $T = 50$ K for LSCO (circles, ref. 13) and YBCO ($p < 0.08$, grey squares, ref. 16). For YBCO at $p > 0.15$ (red squares), we use R_H at $H = 80$ T from **a**. The white diamond (with its error bar) is obtained from the $T = 0$ limit of $R_H(T)$ in strongly overdoped Tl-2201 (ref. 12). The solid black line is a guide to the eye. The red line is $n_H = p$; the blue line is $n_H = 1 + p$. The region where Fermi-surface reconstruction due to CDW order occurs in YBCO is marked as a green band; in that band, $R_H < 0$ at $T \rightarrow 0$. The error bars ($\pm 15\%$) for our four samples (red squares) reflect the uncertainty

$k = (0, \pm\pi)$ and $(\pm\pi, 0)$, the electronic states at the Fermi level must lie near $k = (\pm\pi/2, \pm\pi/2)$, where the four nodes of the d -wave superconducting gap are located. This is indeed what is observed, in the form of nodal Fermi arcs, for example by ARPES (angle-resolved photoemission spectroscopy) in YBCO (ref. 21) and by scanning tunnelling microscopy in Bi-2212 (ref. 8), below $p \approx 0.2$. Given that the relation $n_H = p$ extends down to the lowest dopings (Fig. 4b), two scenarios for these nodal states come to mind. One is associated with the antiferromagnetic order, the other is associated with the Mott insulator.

In the first scenario, antiferromagnetic order with a commensurate wavevector $Q = (\pi, \pi)$ —the order that prevails in YBCO below $p = 0.05$ (Fig. 1a)—would reconstruct the large Fermi surface into four small hole-like nodal pockets whose total volume would contain p carriers, so that $n_H = p$ (see sketch in Fig. 4b). In electron-doped cuprates, an antiferromagnetic quantum critical point is believed to account for the abrupt drop in carrier density detected in the normal-state Hall coefficient²⁷. The question is whether in YBCO magnetic order—present at low temperature up to $p \approx 0.08$ in zero field²⁸ (Fig. 1b)—could extend up to $p^* = 0.19$ when superconductivity is suppressed by a magnetic field of the order of 100 T. An antiferromagnetic quantum critical point at p^* in YBCO could account for the linear temperature dependence of the resistivity²⁹ and possibly also the divergent effective mass⁹.

In the second scenario, the pseudogap phase is a consequence of strong correlations associated with Mott physics. Numerical solutions of the Hubbard model find nodal Fermi arcs at low doping and intermediate temperatures^{30,31}. It has been argued that at $T \rightarrow 0$, the Fermi surface could in fact consist of four hole-like nodal pockets^{32,33} whose total volume would contain p carriers. These arcs/pockets develop even though translational symmetry is not broken. The question is whether such a Mott-based pseudogap can appear at a doping as high as $p = 0.19$.

Overall, the fact that the normal-state carrier density—measured directly in the archetypal cuprate YBCO at low temperature—drops sharply from $n = 1 + p$ to $n = p$ precisely at p^* reveals a robust and

in measuring the geometric factor (see Methods). In YBCO, the carrier density is given by $n = n_H / (\rho_a / \rho_b)$, where ρ_a / ρ_b is the anisotropy ratio of the in-plane resistivity¹⁶. For our samples, $\rho_a / \rho_b \approx 1.5$ (see Methods and Extended Data Fig. 1), so that $n \approx p$ at $p = 0.16$. With decreasing p , the carrier density is seen to drop rapidly from $1 + p$ to p at $p^* = 0.19 \pm 0.01$ (black dotted line), the critical doping for the onset of the pseudogap in YBCO (ref. 11; Fig. 1). The icons above the figure show a sketch of the normal-state Fermi surface in three of the four doping regions: small nodal hole pockets (red) below $p = 0.08$, where magnetic order prevails at low temperature (Fig. 1); small electron pockets (green) between $p = 0.08$ and $p = 0.16$, where charge order (CDW) prevails at low temperature (Fig. 1b); a single large hole surface (blue) above p^* , where the non-superconducting ground state is a correlated metal (grey region).

fundamental new fact about the pseudogap phase: it causes a transformation of the Fermi surface such that its volume suddenly shrinks by one hole per Cu atom. We expect that a microscopic understanding of this transformation will elucidate the enigmatic behaviour of electrons in cuprate superconductors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 October 2015; accepted 5 January 2016.

Published online 22 February 2016.

1. Norman, M. R., Pines, D. & Kallin, C. The pseudogap: friend or foe of high T_c ? *Adv. Phys.* **54**, 715–733 (2005).
2. Wu, T. et al. Magnetic-field-induced charge-stripe order in the high-temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_y$. *Nature* **477**, 191–194 (2011).
3. Ghiringhelli, G. et al. Long-range incommensurate charge fluctuations in $(\text{Y}, \text{Nd})\text{Ba}_2\text{Cu}_3\text{O}_{6+\delta}$. *Science* **337**, 821–825 (2012).
4. Chang, J. et al. Direct observation of competition between superconductivity and charge density wave order in $\text{YBa}_2\text{Cu}_3\text{O}_{6.67}$. *Nature Phys.* **8**, 871–876 (2012).
5. Doiron-Leyraud, N. et al. Quantum oscillations and the Fermi surface in an underdoped high- T_c superconductor. *Nature* **447**, 565–568 (2007).
6. LeBoeuf, D. et al. Lifshitz critical point in the cuprate superconductor $\text{YBa}_2\text{Cu}_3\text{O}_y$ from high-field Hall effect measurements. *Phys. Rev. B* **83**, 054506 (2011).
7. He, Y. et al. Fermi surface and pseudogap evolution in a cuprate superconductor. *Science* **344**, 608–611 (2014).
8. Fujita, K. et al. Simultaneous transitions in cuprate momentum-space topology and electronic symmetry breaking. *Science* **344**, 612–616 (2014).
9. Ramshaw, B. J. et al. Quasiparticle mass enhancement approaching optimal doping in a high- T_c superconductor. *Science* **348**, 317–320 (2015).
10. Senthil, T. On the mass enhancement near optimal doping in high magnetic fields in the cuprates. Preprint at <http://arXiv.org/abs/1410.2096> (2014).
11. Tallon, J. L. & Loram, J. W. The doping dependence of T^* —what is the real high- T_c phase diagram? *Physica C* **349**, 53–68 (2001).
12. Mackenzie, A. P. et al. Normal-state magnetotransport in superconducting $\text{Ta}_2\text{Ba}_2\text{Cu}_3\text{O}_{6+\delta}$ to millikelvin temperatures. *Phys. Rev. B* **53**, 5848–5855 (1996).
13. Ando, Y. et al. Evolution of the Hall coefficient and the peculiar electronic structure of the cuprate superconductors. *Phys. Rev. Lett.* **92**, 197001 (2004).

14. Vignolle, B. *et al.* Quantum oscillations in an overdoped high- T_c superconductor. *Nature* **455**, 952–955 (2008).
15. Hussey, N. E. Phenomenology of the normal state in-plane transport properties of high- T_c cuprates. *J. Phys. Condens. Matter* **20**, 123201 (2008).
16. Segawa, K. & Ando, Y. Intrinsic Hall response of the CuO₂ planes in a chain-plane composite system of YBa₂Cu₃O_y. *Phys. Rev. B* **69**, 104521 (2004).
17. Balakirev, F. F. *et al.* Quantum phase transition in the magnetic-field-induced normal state of optimum-doped high- T_c cuprate superconductors at low temperatures. *Phys. Rev. Lett.* **102**, 017004 (2009).
18. Yoshida, T. *et al.* Systematic doping evolution of the underlying Fermi surface of La_{2-x}Sr_xCuO₄. *Phys. Rev. B* **74**, 224510 (2006).
19. Croft, T. P., Lester, C., Senn, M. S., Bombardi, A. & Hayden, S. M. Charge density wave fluctuations in La_{2-x}Sr_xCuO₄ and their competition with superconductivity. *Phys. Rev. B* **89**, 224513 (2014).
20. Balakirev, F. F. *et al.* Signature of optimal doping in Hall-effect measurements on a high-temperature superconductor. *Nature* **424**, 912–915 (2003).
21. Fournier, D. *et al.* Loss of nodal quasiparticle integrity in underdoped YBa₂Cu₃O_{6+x}. *Nature Phys.* **6**, 905–911 (2010).
22. Hücker, M. *et al.* Competing charge, spin and superconducting orders in underdoped YBa₂Cu₃O_y. *Phys. Rev. B* **90**, 054514 (2014).
23. Blanco-Canosa, S. *et al.* Resonant X-ray scattering study of charge-density wave correlations in YBa₂Cu₃O_{6+x}. *Phys. Rev. B* **90**, 054513 (2014).
24. Wu, T. *et al.* Emergence of charge order from the vortex state of a high-temperature superconductor. *Nature Commun.* **4**, 2113 (2013).
25. Tallon, J. L. *et al.* Zn-induced T_c reduction in high- T_c superconductors: scattering in the presence of a pseudogap. *Phys. Rev. Lett.* **79**, 5294–5297 (1997).
26. Blanco-Canosa, S. *et al.* Momentum-dependent charge correlations in YBa₂Cu₃O_{6+x} superconductors probed by resonant X-ray scattering: evidence for three competing phases. *Phys. Rev. Lett.* **110**, 187001 (2013).
27. Dagan, Y. *et al.* Evidence for a quantum phase transition in Pr_{2-x}Ce_xCuO_{4-x} from transport measurements. *Phys. Rev. Lett.* **92**, 167001 (2004).
28. Haug, D. *et al.* Neutron scattering study of the magnetic phase diagram of underdoped YBa₂Cu₃O_{6+x}. *New J. Phys.* **12**, 105006 (2010).
29. Taillefer, L. Scattering and pairing in cuprate superconductors. *Annu. Rev. Condens. Matter Phys.* **1**, 51–70 (2010).
30. Sénechal, D. & Tremblay, A.-M. S. Hot spots and pseudogaps for hole- and electron-doped high-temperature superconductors. *Phys. Rev. Lett.* **92**, 126401 (2004).
31. Ferrero, M. *et al.* Pseudogap opening and formation of Fermi arcs as an orbital-selective Mott transition in momentum space. *Phys. Rev. B* **80**, 064501 (2009).
32. Rice, T. M. *et al.* A phenomenological theory of the anomalous pseudogap phase in underdoped cuprates. *Rep. Prog. Phys.* **75**, 016502 (2012).
33. Chowdhury, D. & Sachdev, S. Density-wave instabilities of fractionalized Fermi liquids. *Phys. Rev. B* **90**, 245136 (2014).

Acknowledgements We thank the following for discussions: Y. Ando, D. Basov, G. Boebinger, C. Bourbonnais, J. P. Carbotte, A. Carrington, J. Chang, A. V. Chubukov, J. C. Davis, A. Georges, N. E. Hussey, M.-H. Julien, H.-Y. Kee, S. A. Kivelson, P. A. Lee, T. Loew, G. G. Lonzarich, W. Metzner, A. J. Millis, M. R. Norman, C. Pépin, J. Porras, B. J. Ramshaw, T. M. Rice, G. Rikken, S. Sachdev, D. J. Scalapino, C. B. Taillefer, J. L. Tallon, S. Todadri, and A.-M. S. Tremblay. We also thank P. Frings and J.P. Nicolin for help during the measurements in Toulouse. L.T. thanks the Laboratoire National des Champs Magnétiques Intenses (LNCMI) in Toulouse for their hospitality and LABEX NEXT for their support while this work was being performed. A portion of this work was performed at the LNCMI, which is supported by the French ANR SUPERFIELD, the EMFL, and the LABEX NEXT. R.L., D.A.B. and W.N.H. acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (NSERC). L.T. acknowledges support from the Canadian Institute for Advanced Research (CIFAR) and funding from NSERC, the Fonds de recherche du Québec – Nature et Technologies (FRQNT), the Canada Foundation for Innovation (CFI) and a Canada Research Chair.

Author Contributions S.B., W.T., F.L., B.V., D.V., J.B. and C.P. performed the transport measurements at the LNCMI. S.B. and N.D.-L. performed the transport measurements at Sherbrooke. G.G. performed the calculations of the transport coefficients. D.A.B., W.N.H. and R.L. prepared the YBCO single crystals at UBC. S.B., L.T. and C.P. wrote the manuscript, in consultation with all authors. L.T. and C.P. co-supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.T. (louis.taillefer@usherbrooke.ca) or C.P. (cyril.proust@lncmi.cnrs.fr).

METHODS

Samples. Single crystals of $\text{YBa}_2\text{Cu}_3\text{O}_y$ (YBCO) were obtained by flux growth at UBC (ref. 34). The superconducting transition temperature T_c was determined as the temperature below which the zero-field resistance $R = 0$. The hole doping p is obtained from T_c (ref. 35). In order to access dopings above $p = 0.18$, Ca substitution was used, at the level of 1.4% (giving $p = 0.19$) and 5% (giving $p = 0.205$). The samples are rectangular platelets with six contacts applied in the standard geometry, using diffused gold pads.

Measurement of the longitudinal and transverse resistances. The longitudinal resistance R_{xx} and transverse (Hall) resistance R_{xy} of our YBCO samples were measured in Sherbrooke in steady fields up to 16 T and in Toulouse in pulsed fields up to 88 T, using a dual coil magnet developed at the LNCMI-Toulouse to produce non-destructive magnetic fields up to 90 T. The magnetic field profile is shown in Extended Data Fig. 10.

The pulsed-field measurements were performed using a conventional six-point configuration with a current excitation between 5 mA and 10 mA at a frequency of ~ 60 kHz. A high-speed acquisition system was used to digitize the reference signal (current) and the voltage drop across the sample at a frequency of 500 kHz. The data were post-analysed with software to perform the phase comparison. Data for the rise and fall of the field pulse were in good agreement, thus excluding any heating due to eddy currents. Tests at different frequencies showed excellent reproducibility.

Error bars. Note that the resistance of the samples was small due to their geometric factor and their high conductivity in this doping range — typically a few milliohms in the normal state at high fields. Despite the fact that R_{xy} was obtained by anti-symmetrizing the signals measured for a field parallel and anti-parallel to the c axis, a slight negative slope was observed in the Hall coefficient R_H versus H , similar to that found in prior high-field studies^{17,20}. This slope, which may be intrinsic or not, has no impact on any of our conclusions, since they do not depend on the precise absolute value of R_H . Indeed, our conclusions depend on two results: (1) the temperature dependence of R_H at low T , in a given sample; (2) the doping dependence of R_H at low T , at a given temperature. In both cases, what matters is to measure R_H at the same value of H , namely $H = 80$ T. So in Fig. 3c, and Extended Data Figs 7 and 8, where we compare the detailed temperature dependence of $R_H(T)$ in two samples ($p = 0.19$ and $p = 0.205$), the relevant uncertainty is the relative error bar associated with a change of temperature in one sample. That error is defined as the standard deviation of the value of R_H at $H = 80$ T given by the linear fit in Extended Data Fig. 8. The maximum such error bar is shown in Fig. 3 for each of our four samples.

In Fig. 4a, we simply compare the magnitude of R_H in our four samples when measured at $H = 80$ T and $T = 50$ K. As can be seen from the raw data, the negative slope of R_H versus H does not really affect this comparison. What is involved is the error bar on the absolute value of R_H (in $\text{mm}^3 \text{C}^{-1}$), which involves geometric factors and which we estimate to be at most $\pm 15\%$. This error bar is shown in Fig. 4b. Note the excellent quantitative agreement between our data and the data of ref. 16 at $p = 0.16$ and 0.177 (Fig. 3b).

Sample size. No statistical methods were used to predetermine sample size.

Relation between Hall number and carrier density in YBCO. In Fig. 4b, we plot the Hall number $n_H = V/(eR_H)$. In YBCO, the relation between n_H and the carrier density n involves a correction factor, the in-plane anisotropy of transport, so that $n = n_H/(\rho_a/\rho_b)$, where ρ_a and ρ_b are the resistivities along the a and b directions of the orthorhombic structure, respectively. This is because the conducting CuO chains that run along the b axis short-circuit the transverse (Hall) voltage when a current is sent along the a axis¹⁶.

In Extended Data Fig. 1, we show the chain resistivity ρ_{chain} of our YBCO sample at $p = 0.177$, defined as $\rho_{\text{chain}} = 1/(1/\rho_b - 1/\rho_a)$. It displays the known T^2 behaviour of chain conduction³⁶, with $\rho_{\text{chain}} = 50 \mu\Omega \text{ cm}$ at $T = 50$ K. Combined with the $\rho_a(T)$ data plotted in Extended Data Fig. 9, where $\rho_a = 25 \mu\Omega \text{ cm}$ at $T = 50$ K, we get $\rho_a/\rho_b = 1.5$ at $T = 50$ K. We expect a similar anisotropy for the four samples used in our study.

Therefore, if in Fig. 4b we wanted to plot n instead of n_H , we would need to divide n_H by 1.5. The red squares at $p = 0.16$, 0.177 and 0.19 would move down by a factor of 1.5. For $p = 0.16$, this means that $n \approx p$, since $n_H \approx 0.24$. So our claim that $n \approx p$ below p^* remains correct. For $p = 0.205$, we get $n \approx 0.9$, significantly below $1 + p = 1.205$. However, at $p = 0.205$, the value of n_H at $T = 0$ is larger than at $T = 50$ K (Fig. 3c), by a factor of 1.3 or so, giving $n(T \rightarrow 0) \approx 1.2$.

Calculation of the Hall coefficient and resistivity in cuprates. Assuming a single large hole-like Fermi surface, as measured in strongly overdoped Tl-2201, Hussey has shown that one can calculate the resistivity and Hall coefficient using the Jones-Zener expansion³⁷. The model calculates directly the longitudinal and transverse electrical conductivities σ_{xx} and σ_{xy} :

$$\sigma_{xx} = \frac{e^2}{4\pi^3\hbar} \left(\frac{2\pi}{d} \right)^4 \int_0^{\pi/2} \frac{k_F v_F \cos^2(\varphi - \gamma)}{\Gamma \cos \gamma} d\varphi$$

$$\sigma_{xy} = \frac{-e^3 H}{4\pi^3\hbar^2} \left(\frac{2\pi}{d} \right)^4 \int_0^{\pi/2} \frac{v_F \cos(\varphi - \gamma)}{\Gamma} \frac{\partial}{\partial \varphi} \left(\frac{v_F \sin(\varphi - \gamma)}{\Gamma} \right) d\varphi$$

Therefore:

$$R_H = \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} \frac{1}{H}$$

$$\rho_{xx} = \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2}$$

with e the electron charge, \hbar the reduced Planck constant, d the distance between two CuO_2 planes, k_F the Fermi momentum, v_F the Fermi velocity, φ the angle between the momentum \mathbf{k} and the k_x axis in the first Brillouin zone (FBZ), $\gamma(\varphi) = \tan^{-1} \left[\frac{\partial}{\partial \varphi} (\log k_F(\varphi)) \right]$ and Γ the scattering rate. Here we choose k_F and v_F

not to be φ -dependent, that is, the Fermi surface is a perfect cylinder, implying $\gamma(\varphi) = 0$.

We calculate k_F and v_F from hole doping p and effective mass m^* ($= 4.1 m_e$ from quantum oscillations observed in overdoped Tl-2201 (ref. 14)):

$$n = \frac{1 + p}{a^2}$$

$$k_F = \sqrt{2\pi n}$$

$$v_F = \frac{\hbar k_F}{m^*}$$

where a is the in-plane lattice constant parameter (we neglect the slight orthorhombicity of YBCO), and n is the carrier density per CuO_2 plane.

Scenario of inelastic scattering applied to YBCO. Here we discuss the possibility that R_H in YBCO at low temperature is enhanced not by a loss of carrier density but by an increase in inelastic scattering.

It has been shown that anisotropic inelastic scattering can increase the value of $R_H(T)$ even if the Fermi surface remains a single large isotropic cylinder^{15,37}. This mechanism has been argued to account for the rise in R_H measured in overdoped Tl-2201, as occurs when the doping is decreased from $p = 0.3$ to $p = 0.27$, for example (Extended Data Fig. 7).

Here we use the following inelastic scattering model developed by Hussey¹⁵, where the effective scattering rate is given by:

$$1/\Gamma(T, \varphi) = 1/(\Gamma_0 + \Gamma_1 \cos^2(2\varphi)T + \Gamma_2 T^2) + 1/\Gamma_{\max}$$

where T is temperature, Γ_0 is the elastic rate scattering coefficient, Γ_1 is the T -linear inelastic scattering rate coefficient, Γ_2 is the T^2 scattering rate, and $\Gamma_{\max} = v_F/a$ is the maximum scattering rate limited by the lattice constant a .

Here we use this model to fit our Hall data for YBCO at $p = 0.16$, with Γ_1 and Γ_2 the only free parameters (Γ_0 is chosen so that the calculated value of ρ_{xx} at $T = 0$ agrees with experiment). The resulting fit is shown in Extended Data Fig. 9c (solid red line). The corresponding curve of $\rho_{xx}(T) = \rho_a(T)$ is plotted in Extended Data Fig. 9d (solid red line).

In Extended Data Fig. 9, we show how these calculated curves vary when the strength of inelastic scattering is varied, both for R_H (Extended Data Fig. 9c) and for ρ_a (Extended Data Fig. 9d). The calculated curves may be compared with experimental curves in YBCO, shown in the left panels of Extended Data Fig. 9, namely R_H versus T in Extended Data Fig. 9a and ρ_a versus T in Extended Data Fig. 9b. We see that by choosing a large value of Γ_1 , we can fit the Hall data at $p = 0.16$ quite well. The calculated curve drops precipitously below the lowest experimental data point. Then, the decrease in the overall magnitude of R_H versus T with doping can be mimicked in the calculations by decreasing Γ_1 gradually to zero, at which point R_H becomes constant.

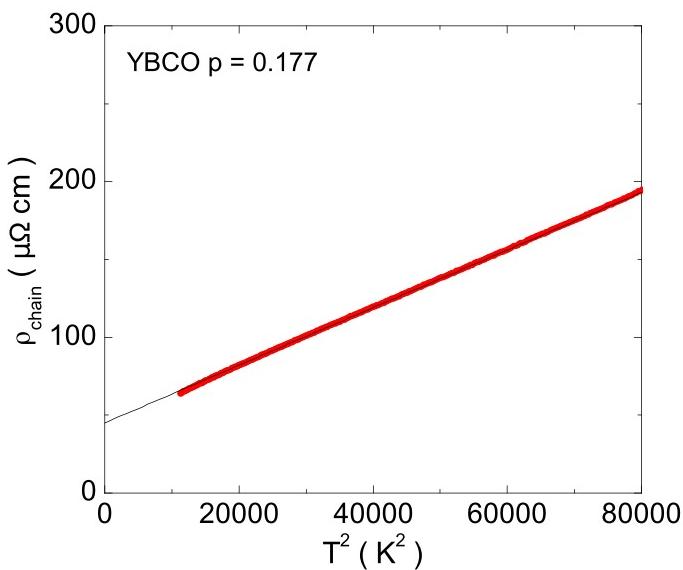
However, while the calculated curves for R_H are consistent with the measured Hall data, the calculated curves for ρ_{xx} are in complete disagreement with the measured ρ_a . This is seen by comparing calculated (Extended Data Fig. 9d) and measured (Extended Data Fig. 9b) values. We see that the tenfold increase in the calculated ρ_{xx} at 50 K, caused by the large increase in Γ_1 , is not at all observed in

the experimental ρ_a , which are essentially independent of doping at $T=50\text{ K}$. In other words, if inelastic scattering were responsible for the increase in R_H at 50 K with underdoping, it would necessarily show up as a comparable (even larger) increase in the resistivity ρ_a at 50 K. The fact that it does not show up in this way rules out inelastic scattering as a mechanism for the sixfold increase in R_H .

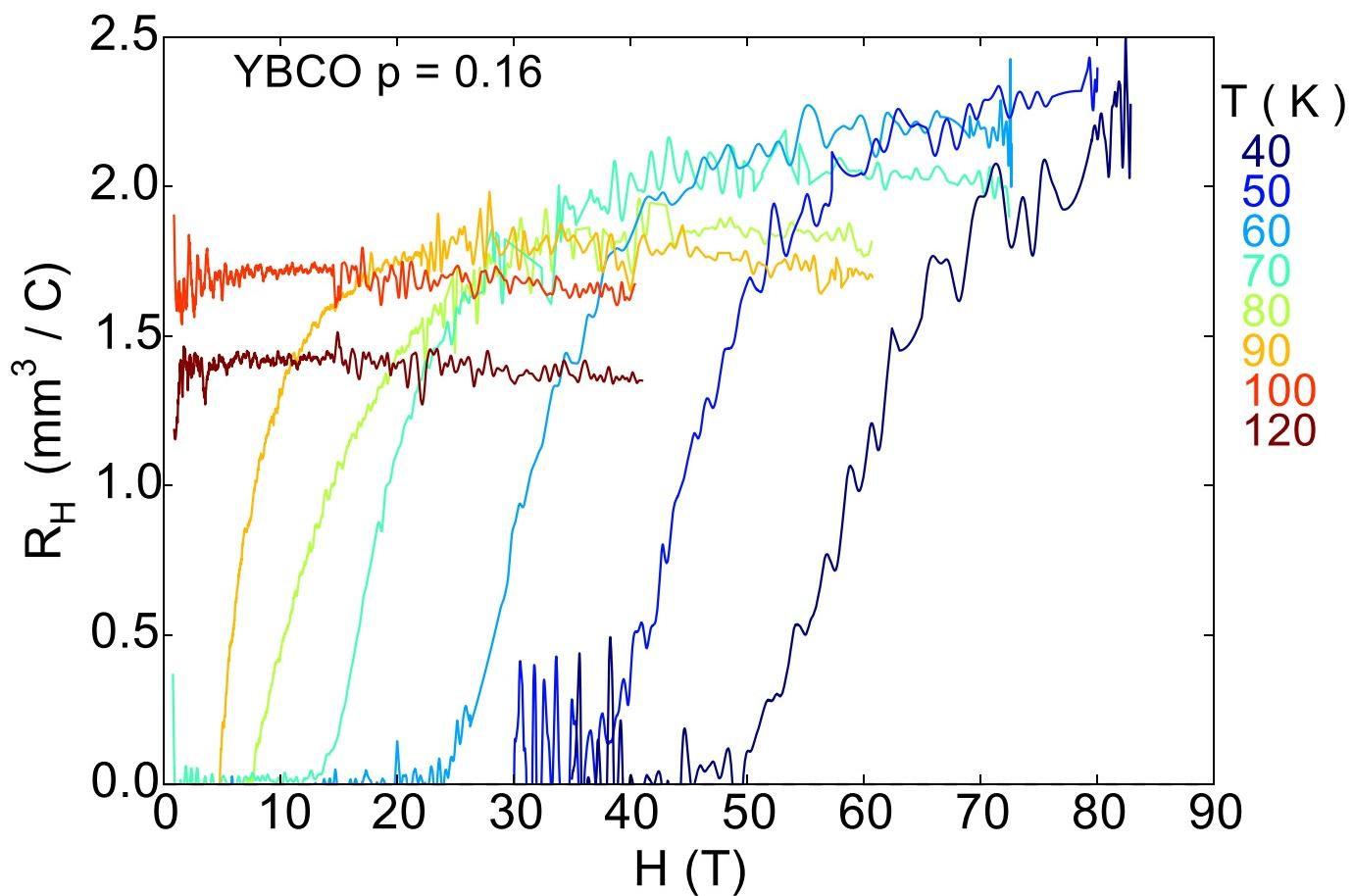
We conclude that the large rise in R_H versus doping is due to a loss of carrier density, and it is a property of the normal-state Fermi surface at $T=0$.

34. Liang, R., Bonn, D. A. & Hardy, W. N. Growth of YBCO single crystals by the self-flux technique. *Phil. Mag.* **92**, 2563–2581 (2012).

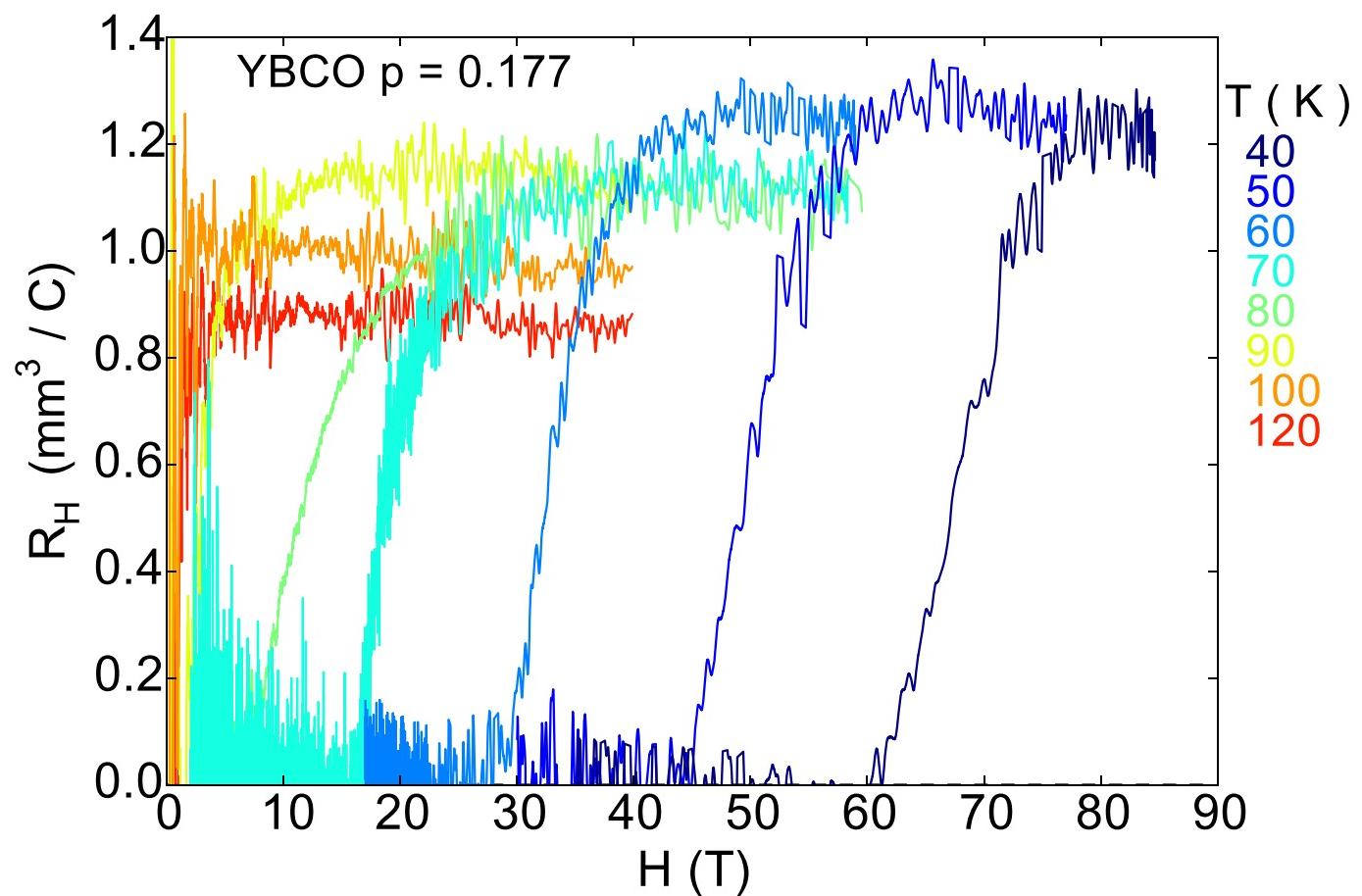
35. Liang, R., Bonn, D. A. & Hardy, W. N. Evaluation of CuO_2 plane hole doping in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ single crystals. *Phys. Rev. B* **73**, 180505 (2006).
36. Gagnon, R., Lupien, C. & Taillefer, L. T^2 dependence of the resistivity in the CuO chains of $\text{YBa}_2\text{Cu}_3\text{O}_{6.9}$. *Phys. Rev. B* **50**, 3458–3461 (1994).
37. Hussey, N. The normal state scattering rate in high- T_c cuprates. *Eur. Phys. J. B* **31**, 495–507 (2003).
38. Manako, T. et al. Transport and structural study of $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ single crystals prepared by the KCl flux method. *Phys. Rev. B* **46**, 11019–11024 (1992).
39. Hussey, N. E. et al. Angular dependence of the c-axis normal-state magnetoresistance in single crystal $\text{Tl}_2\text{Ba}_2\text{CuO}_6$. *Phys. Rev. Lett.* **76**, 122–125 (1996).



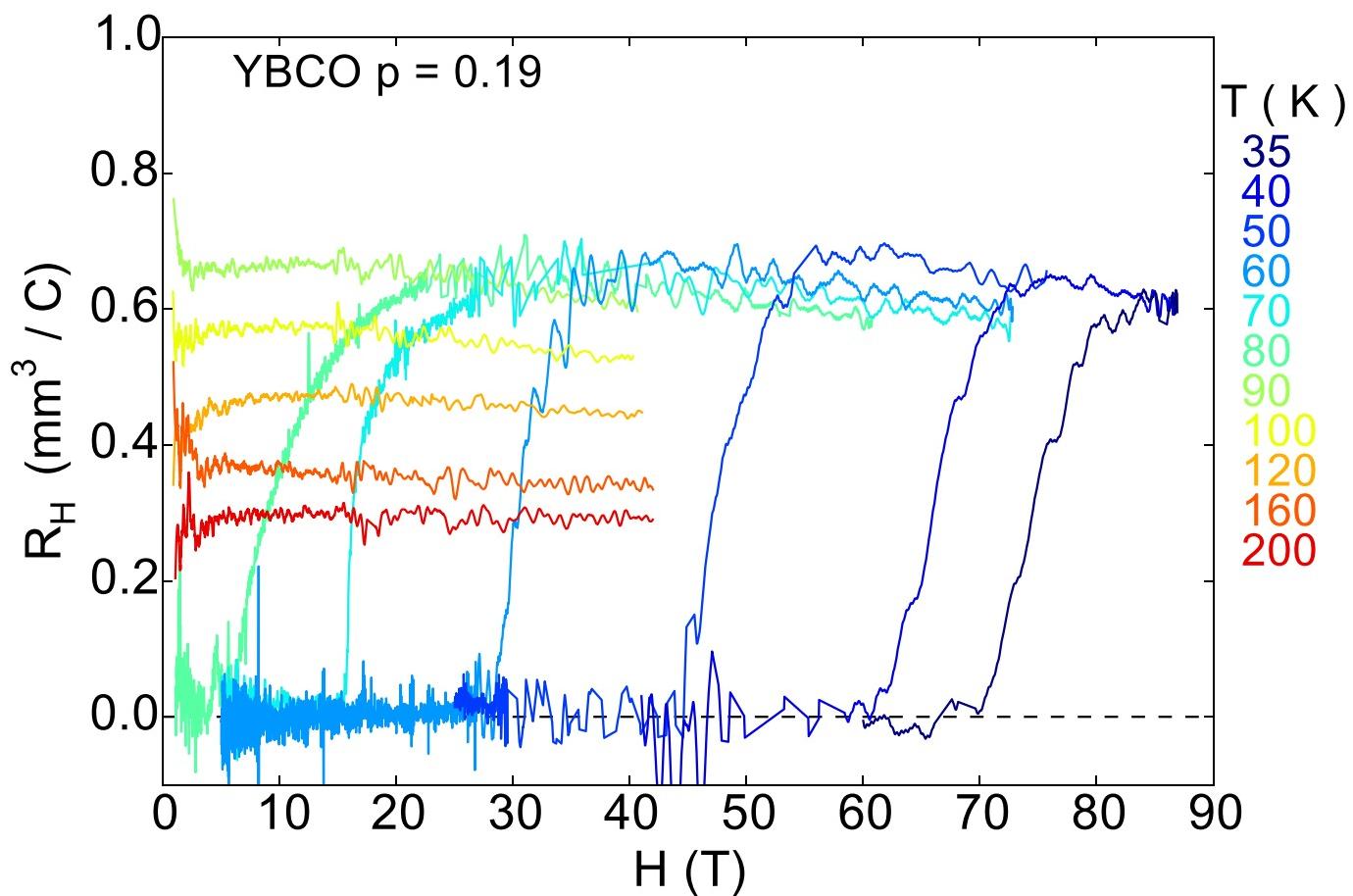
Extended Data Figure 1 | Temperature dependence of resistivity of CuO chains in YBCO at $p = 0.177$. Shown is the chain resistivity in YBCO at $p = 0.177$ (red), defined as $\rho_{\text{chain}} = 1/[(1/\rho_b) - (1/\rho_a)]$, where ρ_a and ρ_b are the in-plane resistivities along the a and b directions of the orthorhombic structure, respectively, plotted versus T^2 . The black line is a linear fit that extrapolates to $\rho_{\text{chain}} = 50 \mu\Omega \text{ cm}$ at $T = 50 \text{ K}$.



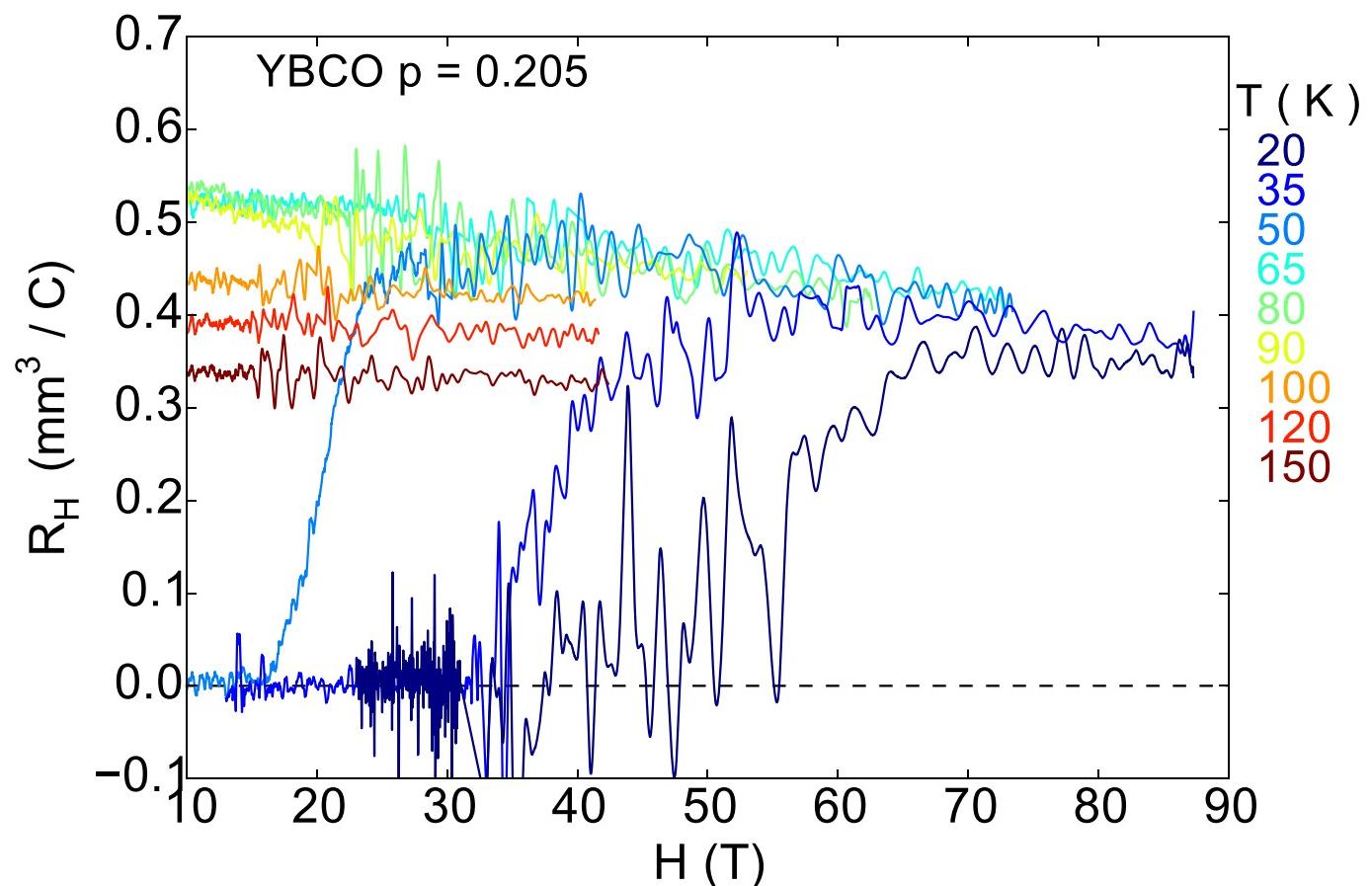
Extended Data Figure 2 | Isotherms of R_H versus H in YBCO at $p = 0.16$. Shown is the magnetic field dependence of the Hall coefficient R_H in our YBCO sample with $y = 6.92$ ($T_c = 93.5$ K; $p = 0.161$) at various temperatures, as indicated (key at right).



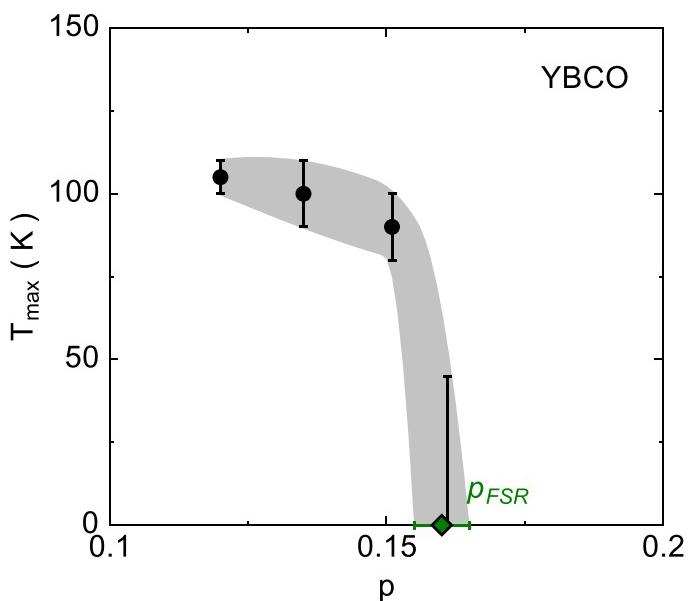
Extended Data Figure 3 | Isotherms of R_H versus H in YBCO at $p = 0.177$. As for Extended Data Fig. 2 but for our YBCO sample with $y = 6.97$ ($T_c = 91$ K; $p = 0.177$).



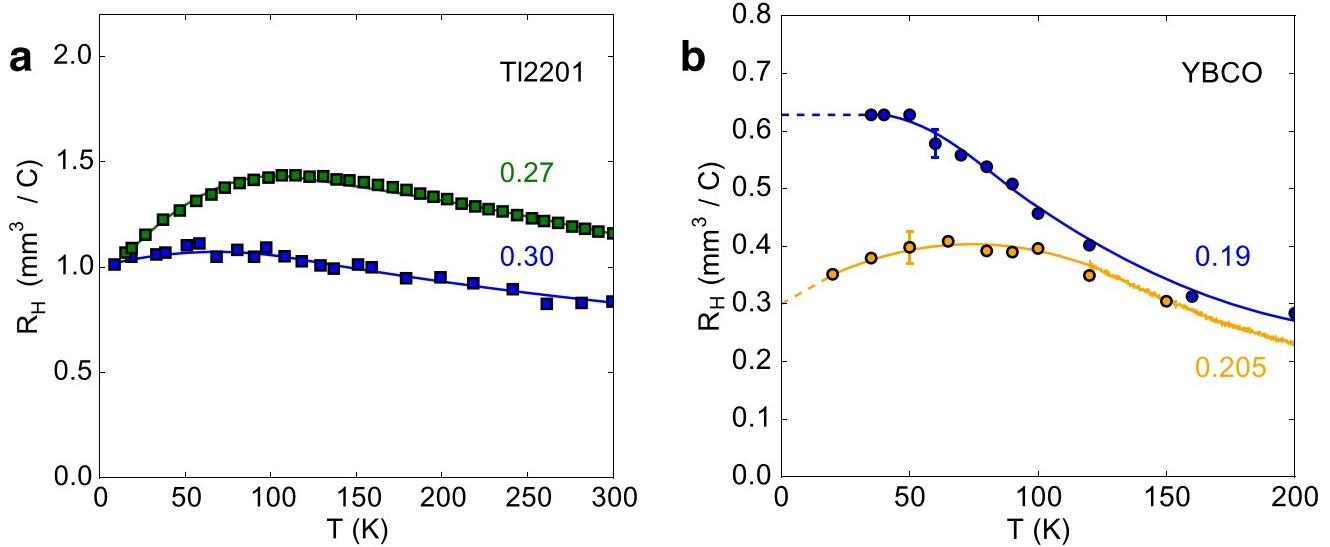
Extended Data Figure 4 | Isotherms of R_H versus H in YBCO at $p=0.19$. As for Extended Data Fig. 2 but for our YBCO sample with $y=6.99$ and 1.4% Ca doping ($T_c=87\text{ K}$; $p=0.19$).



Extended Data Figure 5 | Isotherms of R_H versus H in YBCO at $p = 0.205$. As for Extended Data Fig. 2 but for our YBCO sample with $y = 6.99$ and 5% Ca doping ($T_c = 77$ K; $p = 0.205$).

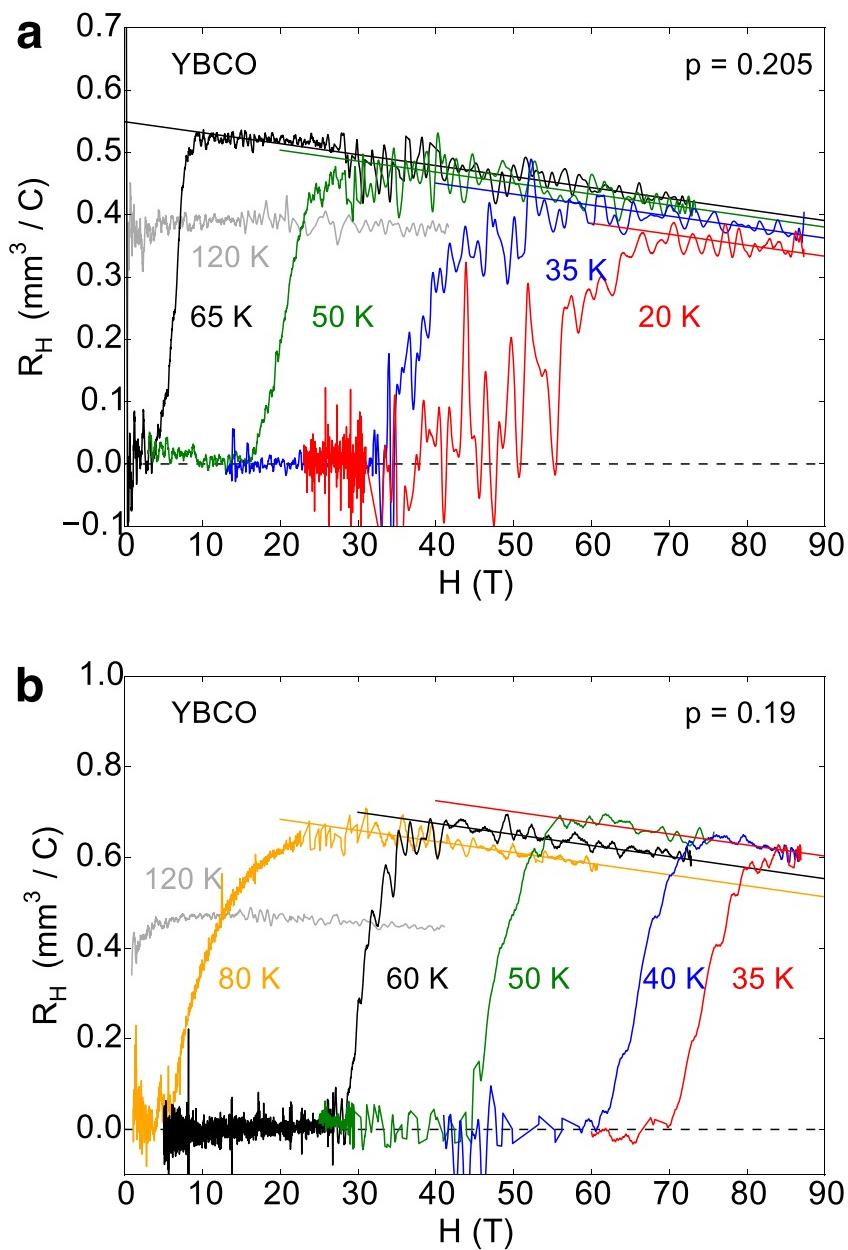


Extended Data Figure 6 | Doping dependence of T_{\max} . Shown is the temperature T_{\max} at which R_H versus T peaks in YBCO (Fig. 3a), plotted versus doping p . At $p = 0.16$, there is no downturn in the normal-state $R_H(T)$ down to 40 K. The $p = 0.16$ data are consistent with $T_{\max} = 0$ (lower bound), with an upper bound at $T_{\max} = 40$ K (shown as black vertical segment). The width of the grey band marks the upper and lower limits for T_{\max} versus p . The green diamond defines the critical doping above which FSR is no longer present, at $p_{FSR} = 0.16 \pm 0.005$, with an error bar defined from the minimal and maximal possible values of T_{\max} . Error bars on the three data points (black dots) represent the uncertainty in defining the peak position of the $R_H(T)$ curves in Fig. 3a.



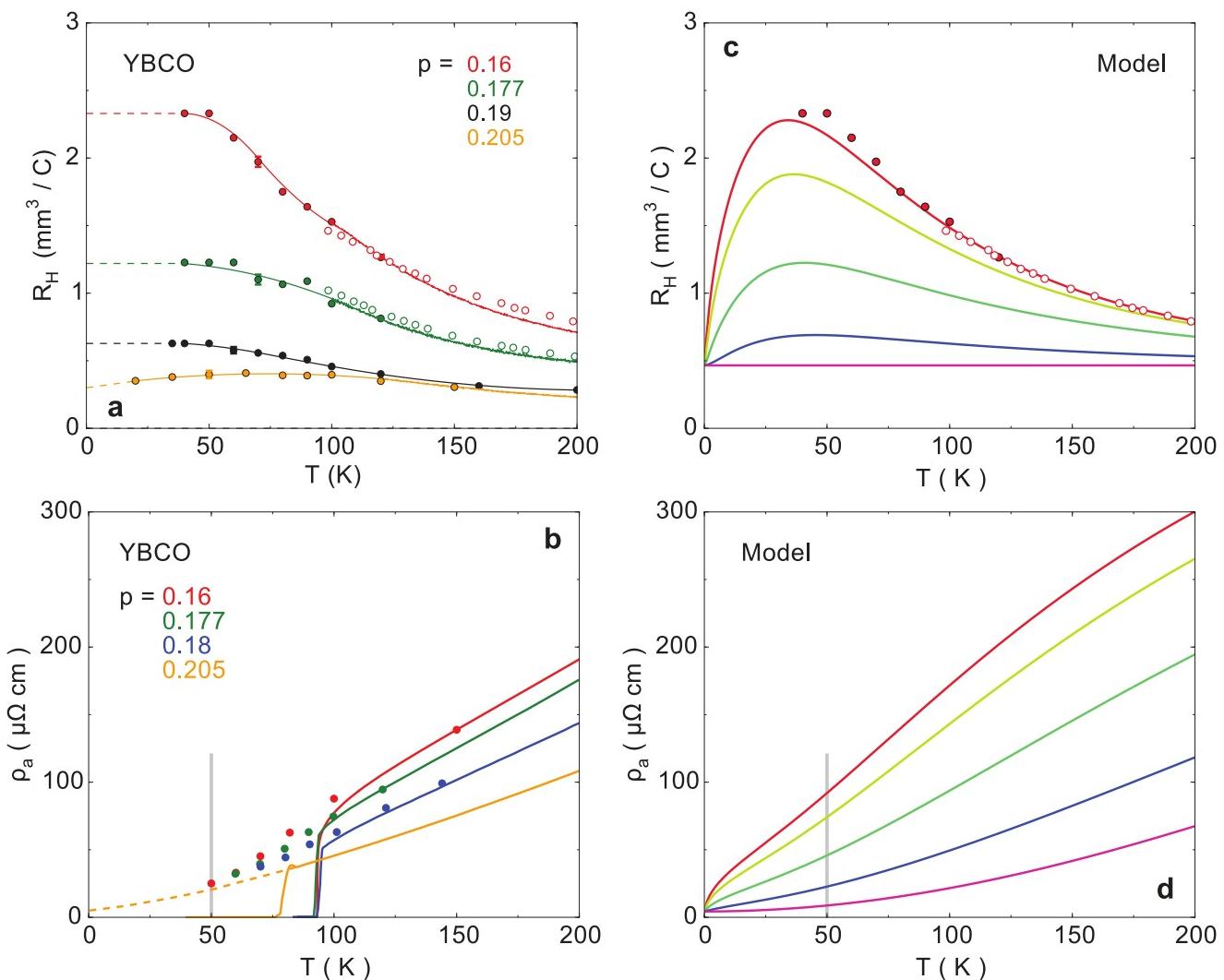
Extended Data Figure 7 | Zoom on R_H versus T in Tl-2201 and YBCO at high doping. **a**, Temperature dependence of R_H in Tl-2201 (squares) at $p = 0.3$ (blue, $T_c = 10$ K; ref. 38) and $p = 0.27$ (green, $T_c = 25$ K; ref. 39). **b**, R_H versus T in YBCO (circles, from Extended Data Figs 4, 5 and 8) at $p = 0.205$ (yellow) and $p = 0.19$ (blue). The dashed lines are an extrapolation of the low- T data to $T = 0$. The YBCO curve at $p = 0.205$ is qualitatively similar to the two Tl-2201 curves, all exhibiting an initial

rise with increasing temperature from $T = 0$, and a characteristic peak at $T \approx 100$ K—two features attributed to inelastic scattering on a large hole-like Fermi surface¹⁵. The YBCO curve at $p = 0.19$ is qualitatively different, showing no sign of a drop at low T (see Extended Data Fig. 8). We attribute the twofold increase in the magnitude of R_H at $T \rightarrow 0$ to a decrease in carrier density as the pseudogap opens at p^* , with p^* located between $p = 0.205$ and $p = 0.19$. The error bars are defined in the legend of Fig. 3.



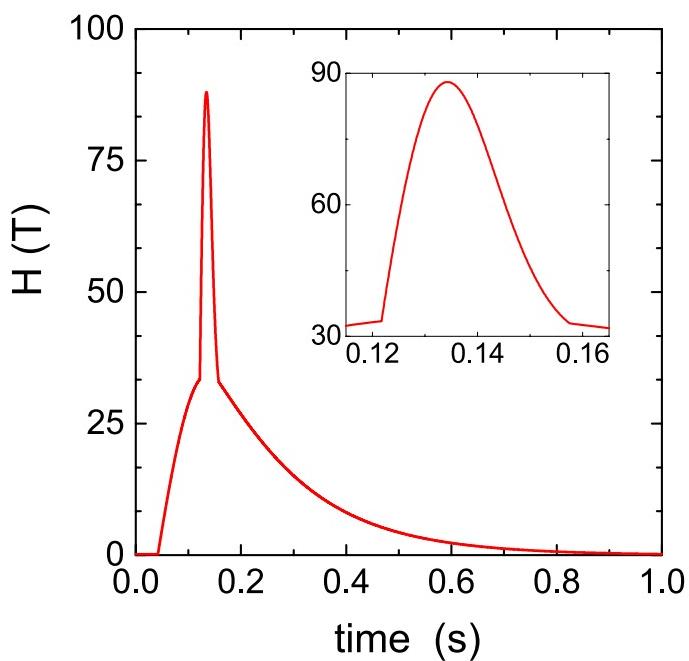
Extended Data Figure 8 | Comparison between $p = 0.205$ and $p = 0.19$.
a, b, The field dependence of the Hall coefficient R_H in YBCO at $p = 0.205$ (a) and $p = 0.19$ (b), for different temperatures as indicated. The colour-coded lines are parallel linear fits to the high-field data. They show that at low temperature R_H decreases upon cooling at $p = 0.205$, while it saturates

at $p = 0.19$. The value of R_H given by the fit line, at $H = 80$ T, is plotted in Fig. 3 and in Extended Data Fig. 7b. Similar fits are used to extract $R_H(80$ T) for $p = 0.16$ and $p = 0.177$ (from data in Extended Data Figs 2 and 3).



Extended Data Figure 9 | Scenario of inelastic scattering. **a**, R_H versus T in YBCO at four dopings p , as indicated (Fig. 3b). **b**, Electrical resistivity ρ_a versus T in YBCO at four dopings, as indicated. Lines are at $H = 0$; dots are in the normal state at high field. **c**, R_H versus T calculated for five values of inelastic scattering, with $\Gamma_1 = 0, 1, 5, 15$ and 25 THz K^{-1} , showing that $R_H(T)$ grows with increasing Γ_1 (see Methods). Dots are from **a**.

d, Corresponding calculated values of the electrical resistivity ρ_a , plotted versus T , using the same parameters and values of Γ_1 as for the colour-coded curves of **c**. The vertical grey lines mark $T = 50$ K, the temperature at which we see a sixfold increase in R_H (**a**), yet no increase in ρ_a (**b**). The calculation can reproduce the large increase in R_H (**c**), but it is accompanied by a tenfold increase in ρ_a (**d**).



Extended Data Figure 10 | Magnetic field profile. Time dependence of the magnetic field pulse in the 90 T dual-coil magnet at the LNCMI in Toulouse. Inset, zoom around maximum field.

Carbon dioxide utilization via carbonate-promoted C–H carboxylation

Aanindeeta Banerjee¹, Graham R. Dick¹, Tatsuhiko Yoshino^{1†} & Matthew W. Kanan¹

Using carbon dioxide (CO_2) as a feedstock for commodity synthesis is an attractive means of reducing greenhouse gas emissions and a possible stepping-stone towards renewable synthetic fuels^{1,2}. A major impediment to synthesizing compounds from CO_2 is the difficulty of forming carbon–carbon (C–C) bonds efficiently: although CO_2 reacts readily with carbon-centred nucleophiles, generating these intermediates requires high-energy reagents (such as highly reducing metals or strong organic bases), carbon–heteroatom bonds or relatively acidic carbon–hydrogen (C–H) bonds^{3–5}. These requirements negate the environmental benefit of using CO_2 as a substrate and limit the chemistry to low-volume targets. Here we show that intermediate-temperature (200 to 350 degrees Celsius) molten salts containing caesium or potassium cations enable carbonate ions (CO_3^{2-}) to deprotonate very weakly acidic C–H bonds ($\text{pK}_a > 40$), generating carbon-centred nucleophiles that react with CO_2 to form carboxylates. To illustrate a potential application, we use C–H carboxylation followed by protonation to convert 2-furoic acid into furan-2,5-dicarboxylic acid (FDCA)—a highly desirable bio-based feedstock⁶ with numerous applications, including the synthesis of polyethylene furandicarboxylate (PEF), which is a potential large-scale substitute for petroleum-derived polyethylene terephthalate (PET)^{7,8}. Since 2-furoic acid can readily be made from lignocellulose⁹, CO_3^{2-} -promoted C–H carboxylation thus reveals a way to transform inedible biomass and CO_2 into a valuable feedstock chemical. Our results provide a new strategy for using CO_2 in the synthesis of multi-carbon compounds.

The chemistry described here is inspired by ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which effects C–C bond formation in the Calvin cycle by deprotonating a C–H bond of ribulose-1,5-bisphosphate and exposing the resulting carbon-centred nucleophile to CO_2 to form a carboxylate ($\text{C}-\text{CO}_2^-$)¹⁰. Emulating this strategy synthetically requires deprotonating un-activated C–H bonds using a simple base that does not have a large CO_2 footprint. To meet these requirements, we envisioned a CO_3^{2-} -promoted C–H carboxylation reaction, wherein CO_3^{2-} reversibly deprotonates a C–H bond to generate HCO_3^- and a carbon-centred nucleophile that reacts with CO_2 to form $\text{C}-\text{CO}_2^-$ (Fig. 1a). HCO_3^- decomposition results in a net consumption of one-half equivalents of CO_3^{2-} and CO_2 per $\text{C}-\text{CO}_2^-$ produced. The cycle could be closed by protonating $\text{C}-\text{CO}_2^-$ with strong acid and using electrodialysis to regenerate the acid and base^{11,12}, effecting a net transformation of C–H and CO_2 into $\text{C}-\text{CO}_2\text{H}$ without using any other stoichiometric reagents. Alternatively, CO_2 -promoted esterification could be used to convert the carboxylate into an ester ($\text{C}-\text{CO}_2\text{R}$) and regenerate CO_3^{2-} directly¹³. Previously, researchers have shown that Cs_2CO_3 can deprotonate alkynyl¹⁴, allylic¹⁵, and activated heteroaryl C–H bonds with pK_a values of up to 27 (ref. 16) in organic solvents at elevated temperature (where pK_a is the negative base-10 logarithm of the acid dissociation constant). However, using CO_3^{2-} -promoted C–H carboxylation for commodity synthesis requires deprotonating C–H bonds that are considerably less acidic.

Carbonate-promoted C–H carboxylation is particularly desirable for the synthesis of polymer units from biomass. A longstanding goal of renewable polymer chemistry is a scalable synthesis of FDCA from lignocellulose to replace petroleum-derived terephthalic acid for polyester synthesis (Fig. 1b)⁶. In particular, PEF has been reported to have superior physical properties to PET, a polymer produced at 50 megatons a year as an industrial commodity⁸. Current approaches to FDCA synthesis use dehydration processes to convert hexose sugars into hydroxymethyl furfural (HMF), which is then oxidized to form FDCA¹⁷. Recent work has improved the efficiency of converting fructose to HMF^{8,18} and it has been estimated that PEF production from fructose would emit 50% less CO_2 than the established process for PET production⁷. However, producing FDCA on a scale commensurate with terephthalic acid and achieving maximal reduction in CO_2 emissions will require using lignocellulose as the feedstock. Converting lignocellulose into HMF is very challenging because the hexoses are incorporated into intractable cellulose fibres¹⁷. Researchers have recently reported a high-yielding conversion of lignocellulose to HMF in ionic liquids^{19,20} and an efficient process that converts lignocellulose into sugar monomers²¹. Despite these advances, an economical, large-scale lignocellulose-to-HMF process has not been demonstrated^{17,22}. In contrast, the conversion of lignocellulose to furfural is a decades-old industrial process that is currently performed on a scale of about 400 kilotons a year^{9,22}. Furthermore, several methods are available for oxidizing furfural to 2-furoic acid^{23,24}. At present, however, the available methods for converting 2-furoic acid into FDCA are inefficient, unselective, and consume stoichiometric amounts of energy-intensive reagents^{25,26}. CO_3^{2-} -promoted C–H carboxylation could be used to convert 2-furoic acid into FDCA and thereby open a new route to PEF using a lignocellulose-derived monomer that is already produced industrially.

The CO_3^{2-} -promoted C–H carboxylation reaction required for FDCA synthesis is the conversion of furan-2-carboxylate into furan-2,5-dicarboxylate (FDCA²⁻) (Fig. 2a). Assuming it is similar to an un-substituted furan²⁷, the pK_a of the C–H at the 5 position of furan-2-carboxylate is ~ 35 . Deprotonation of this C–H has previously required lithium diisopropylamide or *n*-butyllithium²⁶. We hypothesized that CO_3^{2-} would deprotonate furan-2-carboxylate if the reaction were performed in a molten salt with a high concentration of alkali cations to stabilize the conjugate base by ion pairing. To test this hypothesis, we attempted C–H carboxylation with mixtures consisting of an alkali metal salt of 2-furoic acid and an alkali metal carbonate. With these components, the reaction was found to proceed efficiently when caesium ions (Cs^+) were used (Fig. 2a and Extended Data Table 1a). When 1 mmol of caesium furan-2-carboxylate and 0.55 mmol Cs_2CO_3 were heated at 260 °C under a CO_2 flow of 40 ml min⁻¹ in a tube furnace, FDCA²⁻ was formed in 76% yield after 12 h (Extended Data Fig. 1). The mass balance was composed of unreacted starting material and decomposition products including acetate. Reactions performed in a Parr reactor showed improved yields and less decomposition.

¹Department of Chemistry, Stanford University, Stanford, California, USA.

[†]Present address: Faculty of Pharmaceutical Sciences, Hokkaido University, Kita-ku, Sapporo 060-0812, Japan.

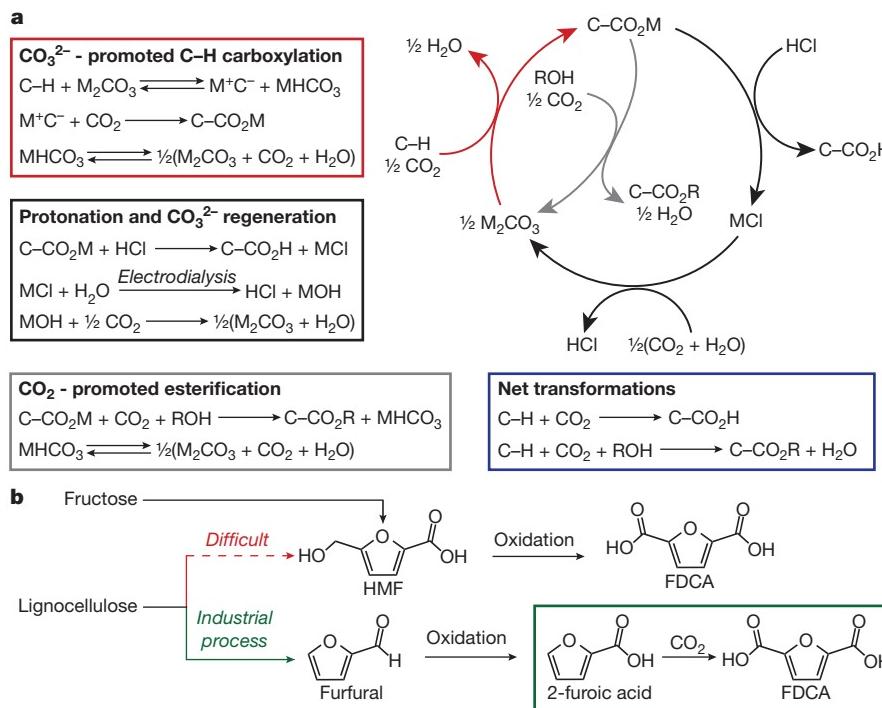
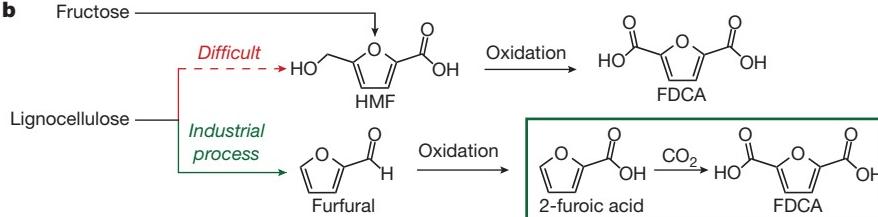
**b**

Figure 1 | CO₂ utilization cycle. **a**, C-H deprotonation by CO₃²⁻ forms a carbon-centred nucleophile (C⁻) that reacts with CO₂ to form a C-C bond (red box and arrows). Protonation and electrodialysis (black box and arrows) produces a carboxylic acid. Alternatively, CO₂-promoted esterification (grey box and arrows) produces an ester. Net transformations

In 1-mmol-scale reactions at 200 °C under a pressure of 8 bar CO₂, FDCA²⁻ was formed in a 77% yield after 2 h and 89% yield after 5 h, with only 5% decomposition products. In 10-mmol-scale reactions under similar conditions, FDCA²⁻ was formed in a 78% yield after 5 h and 81% yield after 10 h (Extended Data Fig. 3). Further increasing the reaction time did not appreciably increase the FDCA²⁻ yield, while increasing the temperature diminished the yield because of increased decomposition. Increasing the CO₂ pressure slowed the reaction by sequestering CO₃²⁻ in the form of HCO₃⁻. Finally, a 100-mmol-scale reaction was performed under ~1 bar CO₂ in a rotating round-bottom flask in a 260 °C bath. After 48 h, FDCA²⁻ was formed in a 71% yield. The scaling behaviour suggests that the reaction takes place at the molten salt–CO₂ interface. The reaction slows and the yield decreases somewhat as the scale is increased because the surface area-to-volume ratio decreases. Improved yields and rates are anticipated with reactors that disperse the salt more effectively.

No FDCA²⁻ was observed when carboxylation reactions were attempted using mixtures of furan-2-carboxylate and CO₃²⁻ with alkali cations other than Cs⁺. These salt mixtures required temperatures much higher than 200 °C to melt, which resulted in decomposition reactions. The restriction to Cs⁺ can be lifted, however, by incorporating another carboxylate salt to attain a semi-molten solution. For example, heating potassium furan-2-carboxylate with 0.5 equivalents of K₂CO₃ and 1 equivalent of potassium isobutyrate at 320 °C under 40 ml min⁻¹ CO₂ resulted in 62% potassium FDCA²⁻ (Extended Data Fig. 5a). Similar results were obtained with potassium acetate as an additive. Thus, C-H carboxylation does not require Cs⁺ *per se*, but caesium salts do typically have lower melting points.

While the carboxylation results are consistent with the mechanistic scheme in Fig. 1a, there are other possible mechanisms that do not involve a carbanion intermediate. To probe the acid–base properties of furan-2-carboxylate independently, an isotope exchange experiment was performed between furan-2-carboxylate and acetate (Fig. 2c). (The pK_a of acetate is >33; ref. 28). A mixture of 1 mmol

are shown in the blue box. **b**, Potential application for FDCA synthesis from lignocellulose. Current approaches rely on converting lignocellulose into HMF, a difficult reaction that has not been commercialized. C-H carboxylation enables the use of furfural, which has been produced industrially from lignocellulose for many decades. M, metal; R, alkyl group.

caesium furan-2-carboxylate, 1 mmol CD₃CO₂Cs, and 1.1 mmol Cs₂CO₃ was heated under N₂ in the Parr reactor to 200 °C for 1 h. ¹H, ²H, and ¹³C nuclear magnetic resonance (NMR) spectra of the crude product mixture showed hydrogen/deuterium (H/D) scrambling between acetate and the 5-position of furan-2-carboxylate and, to a lesser extent, the 3 and 4 positions (Extended Data Figs 6 and 7). The H content remaining in furan-2-carboxylate indicated that the exchange was ~60% complete. High-resolution mass spectrometry showed the fully protonated, and singly, doubly, and triply deuterated furan-2-carboxylate (Fig. 2d). When a 1:1 mixture of caesium furan-2-carboxylate and CD₃CO₂Cs was heated in the absence of Cs₂CO₃ at 200 °C, ~15% H/D exchange was observed, with nearly all of the exchange occurring at the 5 position (Fig. 2c and Extended Data Fig. 8). Thus, at 200 °C in a molten salt, a carboxylate is able to deprotonate the 5 position of furan-2-carboxylate, and CO₃²⁻ is able to deprotonate all positions. The selectivity seen in the carboxylation reaction suggests a greater abundance of the carbanion that leads to FDCA²⁻.

Additional substrates were evaluated to gain further insight into the CO₃²⁻-promoted C-H carboxylation reaction. Heating caesium thiophene-2-carboxylate with 0.55 equivalents of Cs₂CO₃ under flowing CO₂ at 325 °C resulted in a 71% yield of thiophene-2,5-dicarboxylate after 4 h (Extended Data Fig. 2). This substrate required a higher temperature than furan-2-carboxylate to form a semi-molten solution. To see whether C-H carboxylation is possible for substantially weaker acids, we evaluated the reactivity of benzoate. A previous study reported the carboxylation of caesium benzoate with Cs₂CO₃ at extreme CO₂ pressure (400 bar) at 380 °C via electrophilic activation of the aryl ring with a CO₃²⁻–CO₂ complex²⁹. The results with furan-2-carboxylate indicate that benzoate deprotonation by CO₃²⁻ is feasible, which would enable carboxylation under milder conditions at low CO₂ pressures. Remarkably, heating caesium benzoate with 0.55 equivalents of Cs₂CO₃ to 320 °C under 8 bar CO₂ resulted in a combined yield of 66% for a mixture of phthalates and tri- and tetra-carboxylates (Fig. 3a and Extended Data Fig. 4). The ability of CO₃²⁻

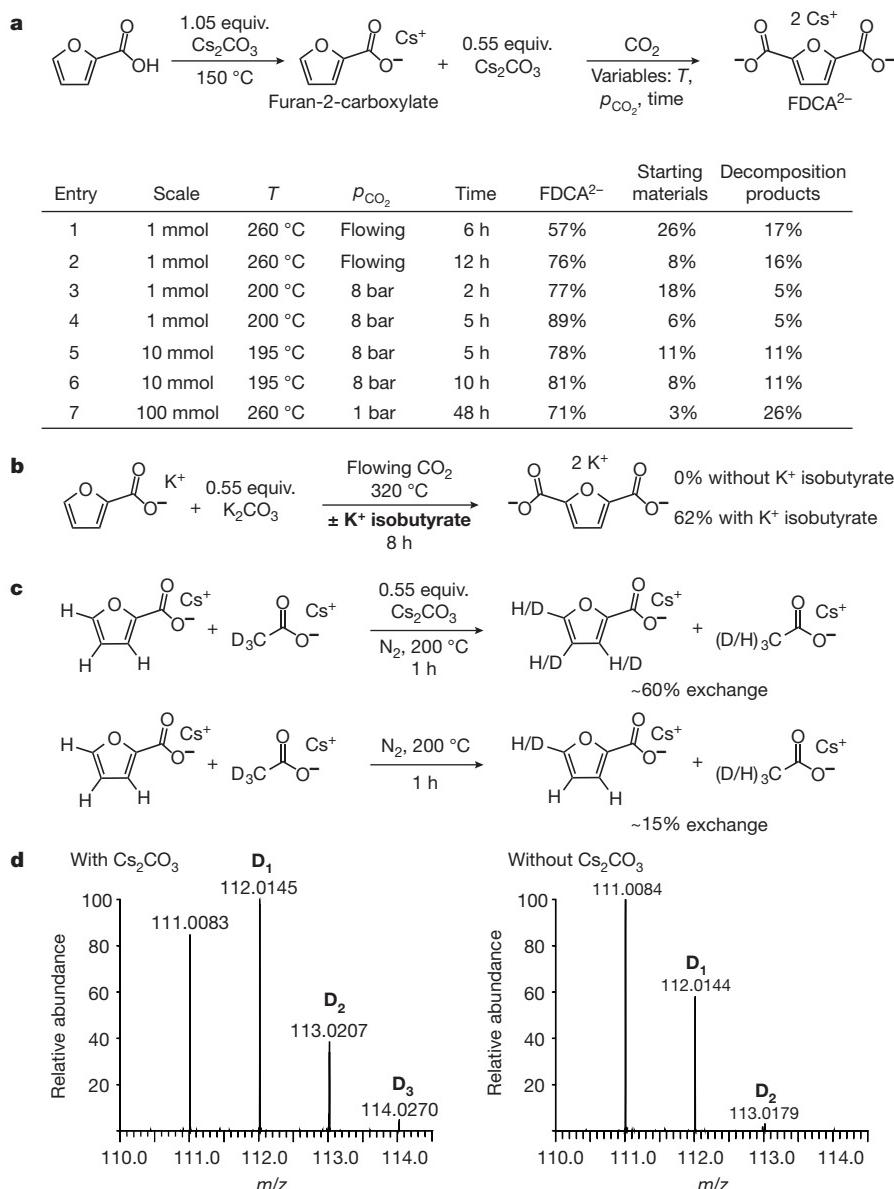


Figure 2 | C–H carboxylation of furan-2-carboxylate. **a**, Carboxylation of caesium-furan-2-carboxylate to caesium FDCA $^{2-}$. p_{CO_2} is the pressure of carbon dioxide at the reaction temperature T. 1.25 equivalents of Cs₂CO₃ were used for entry 7. **b**, Carboxylation of potassium furan-2-carboxylate to potassium FDCA $^{2-}$ using potassium isobutyrate as a low-melting co-salt. **c**, Evidence for C–H deprotonation. Cs₂CO₃ catalyses isotopic

exchange between deuterated acetate and all positions of furan-2-carboxylate. In the absence of Cs₂CO₃, partial exchange is observed at the 5 position. The extent of exchange is calculated based on the integration of the ¹H NMR (see Extended Data Figs 6b and 8a). **d**, Mass spectrometry of the exchange reactions in c. D₁, D₂, and D₃ correspond to singly, doubly and triply deuterated furan-2-carboxylate. m/z, mass to charge ratio.

to deprotonate the C–H bonds of a phenyl ring was tested independently by heating a mixture of ¹²C₆D₅CO₂Cs, ¹³C₆H₅CO₂Cs, and 0.55 equivalents of Cs₂CO₃ to 320 °C under N₂ for 30 min. ¹H NMR analysis of the products revealed H/D scrambling at all positions on the benzoate ring (Fig. 3b). No H/D exchange was observed in the absence of Cs₂CO₃ (Extended Data Fig. 9).

The results with benzoate suggest that benzene would undergo C–H carboxylation if exposed to CO₃ $^{2-}$ in a molten salt. Heating Cs₂CO₃ under benzene and CO₂ at temperatures up to 380 °C resulted in no reaction. The lack of reactivity can be attributed to the fact that Cs₂CO₃ does not melt. To provide a molten component, reactions were performed in the presence of caesium isobutyrate. Heating 1.5 mmol Cs₂CO₃ and 1 mmol caesium isobutyrate to 340–380 °C under 31 bar CO₂ and 42 bar benzene resulted in the formation of benzoate, phthalates, and benzene tricarboxylates (Fig. 3c and Extended Data Fig. 5b). The amount of benzene carboxylation products corresponded

to 12% of the Cs₂CO₃ at 350 °C, and 19% at 360 °C. In addition to benzene carboxylation, isobutyrate carboxylation to dimethyl malonate and decomposition to formate and acetate also occurred under these conditions (Extended Data Table 1b). The carboxylation of benzene is more challenging than benzoate because there is a larger entropic penalty and the solubility of benzene in the carboxylate salt is likely to be very low. Nevertheless, these results demonstrate that CO₃ $^{2-}$ -promoted hydrocarbon carboxylation is possible.

Scalable CO₃ $^{2-}$ -promoted C–H carboxylation requires facile product isolation and highly efficient recovery of the alkali cation. Researchers have recently reported the conversion of carboxylates to methyl esters by reaction with CO₂ and methanol at 160–200 °C (ref. 13). Similar conditions were found to effect double esterification of FDCA $^{2-}$ to dimethyl furan-2,5-carboxylate (DMFD) in moderate yield (Fig. 4a). Heating 1 mmol of caesium FDCA $^{2-}$ to 200 °C in 100 ml anhydrous methanol under 45 bar CO₂ for 30 min resulted in 50% yield of DMFD,

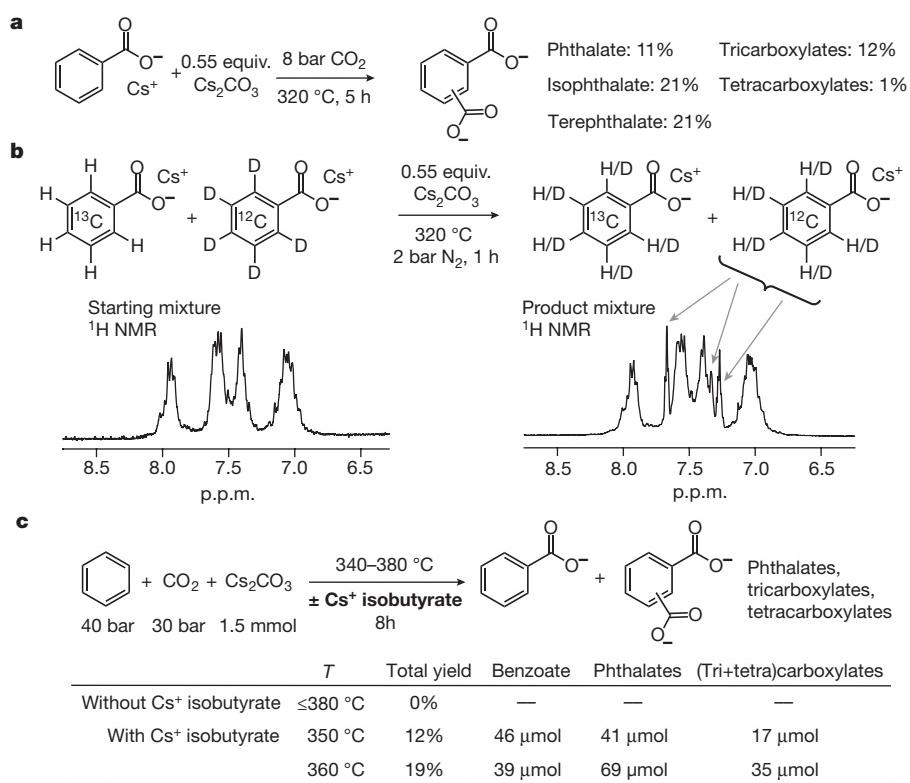


Figure 3 | C-H carboxylation of benzoate and benzene. **a**, Carboxylation of caesium benzoate; yields are given with respect to benzoate. **b**, CO_3^{2-} -catalysed H/D isotope exchange between differentially labelled benzoates. Exchange is seen in the appearance of ^1H peaks associated with $^{12}\text{C}-\text{H}$ bonds. No exchange is observed in the absence of Cs_2CO_3 . **c**, Carboxylation of benzene in the presence or absence of caesium isobutyrate. The yield refers to the amount of CO_3^{2-} consumed for the formation of benzene carboxylation products. equiv., equivalents; p.p.m., parts per million.

36% of the mono-ester, 5-(methoxycarbonyl)furan-2-carboxylate (MMFD), and 14% remaining FDCA^{2-} . Subjecting a 1:1 mixture of DMFD and Cs_2CO_3 to the same conditions resulted in a similar product mixture, indicating that the reaction is under thermodynamic control. To test whether Cs_2CO_3 can be recovered, sequential

carboxylation/esterification cycles were performed (Fig. 4b). 1 mmol of 2-furoic acid and 1.05 equivalents of Cs_2CO_3 were subjected to a one-pot, two-step sequence of carboxylation and esterification. Extraction of the crude product mixture afforded 388 μmol of DMFD. The residual material containing FDCA^{2-} and Cs_2CO_3 was combined

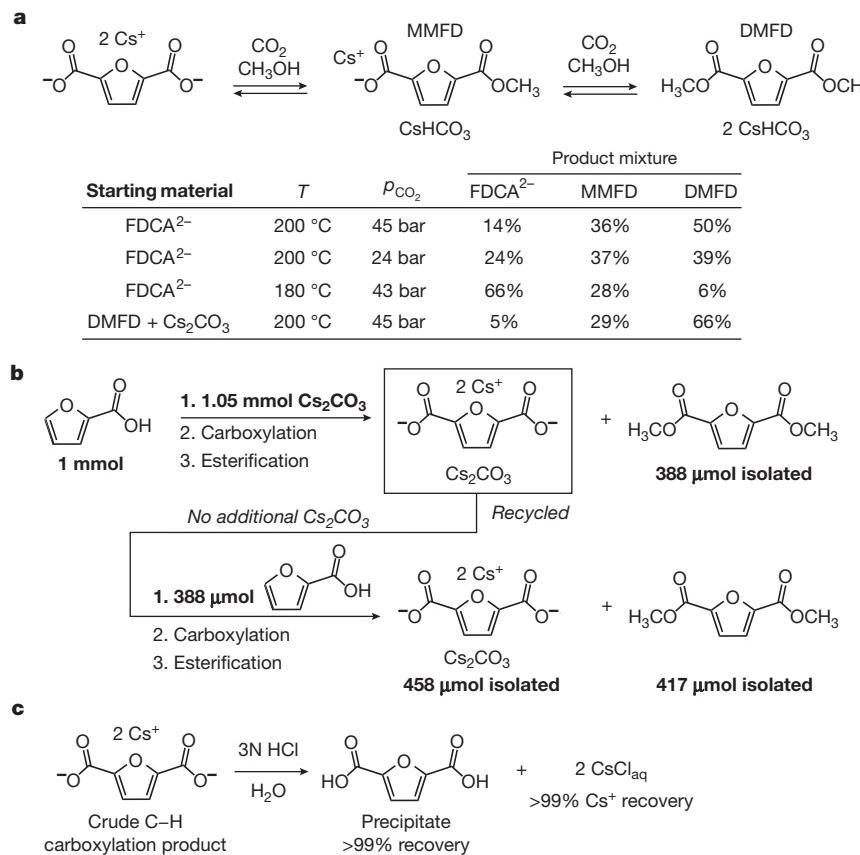


Figure 4 | Product isolation and Cs^+ recovery. **a**, CO_2 promotes the esterification of FDCA^{2-} to MMFD and DMFD. The table shows the results of four 1-mmol-scale reactions performed in 100 ml anhydrous methanol starting with either caesium FDCA^{2-} or DMFD. **b**, Two sequential carboxylation/esterification sequences with Cs_2CO_3 recycling. The total carboxylation yield (91%) is similar to the yield for an individual experiment, indicating complete recovery of Cs_2CO_3 after the first sequence. **c**, Isolation of FDCA by protonation with strong acid. CsCl is retained in the aqueous solution.

with a second aliquot of 388 µmol 2-furoic acid and carried through a second sequence of carboxylation and esterification. Extraction yielded 417 µmol DMFD, while analysis of the unreacted material indicated 458 µmol of a mixture of FDCA²⁻ and MMFD (Extended Data Fig. 10). The total amount of carboxylation products (DMFD + FDCA²⁻ + MMFD) was 1.26 mmol, which is 91% of the total amount of 2-furoic acid substrate (1.39 mmol) used for the two carboxylation/esterification sequences. These results indicate that the Cs₂CO₃ produced from esterification of caesium FDCA²⁻ can be reused in a subsequent C–H carboxylation without loss of yield, which in principle enables a cycle that converts 2-furoic acid into DMFD using only CO₂ and methanol as stoichiometric reagents. It may be possible to improve esterification yields under milder conditions by removing water *in situ*³⁰ or using an alternative solvent.

As an alternative to esterification, treatment of crude caesium FDCA²⁻ from a C–H carboxylation reaction with 3 N HCl affords immediate precipitation of FDCA, leaving CsCl in the aqueous solution with >99% Cs⁺ recovery (Fig. 4c). To complete the cycle, bipolar membrane electrodialysis^{11,12} could be used to convert CsCl into HCl and CsOH solutions. HCl is recycled for the protonation step, while CsOH is reacted with 2-furoic acid and CO₂ to generate the starting material for C–H carboxylation. The energy requirement for converting aqueous alkali chloride solutions into HCl and alkali hydroxide solutions is ~0.08 kWh per mole of alkali chloride^{11,12}, which would correspond to ~1 kWh per kilogram of FDCA. While additional energy would be required for water removal in each cycle, using highly concentrated solutions would minimize this requirement. The overall process would convert 2-furoic acid into FDCA without using any organic solvents or product distillation steps.

Our results demonstrate a very simple strategy for engaging CO₂ in C–C bond formation that does not require synthetic or biological catalysts. The ability to deprotonate unactivated C–H bonds opens the possibility of using this approach to prepare numerous high-volume targets. In particular, combining carboxylation with hydrogenation reactions may enable the synthesis of multi-carbon alcohols and hydrocarbons using CO₂ and renewable H₂.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 September 2015; accepted 29 January 2016.

- Mikkelsen, M., Jorgensen, M. & Krebs, F. C. The teraton challenge. A review of fixation and transformation of carbon dioxide. *Energ. Environ. Sci.* **3**, 43–81 (2010).
- Aresta, M., Dibenedetto, A. & Angelini, A. Catalysis for the valorization of exhaust carbon: from CO₂ to chemicals, materials, and fuels. Technological use of CO₂. *Chem. Rev.* **114**, 1709–1742 (2014).
- Mander, L. N., Adreatta, J. R. & Darenbourg, D. J. Carbon dioxide. *Encyclopedia of Reagents for Organic Synthesis (e-EROS)* <http://onlinelibrary.wiley.com/doi/10.1002/047084289X.rc011.pub2/full> (2008).
- Cai, X. & Xie, B. Direct carboxylative reactions for the transformation of carbon dioxide into carboxylic acids and derivatives. *Synthesis* **45**, 3305–3324 (2013).
- Liu, A. H., Yu, B. & He, L. N. Catalytic conversion of carbon dioxide to carboxylic acid derivatives. *Greenhouse Gas. Sci. Technol.* **5**, 17–33 (2015).
- Werpy, T. et al. Top value added chemicals from biomass. In *Results of Screening for Potential Candidates from Sugars and Synthesis Gas* Vol. 1, 26–28, <http://www.nrel.gov/docs/fy04osti/35523.pdf> (US DOE, 2004).
- Eerhart, A. J. J. E., Faaij, A. P. C. & Patel, M. K. Replacing fossil based PET with biobased PEF; process analysis, energy and GHG balance. *Energ. Environ. Sci.* **5**, 6407–6422 (2012).
- de Jong, E., Dam, M., Sipos, L. & Gruter, G. Furandicarboxylic acid (FDCA), a versatile building block for a very interesting class of polyesters. *Biobased Monomers Polymers Mater.* **1105**, 1–13 (2012).

- Lange, J. P., van der Heide, E., van Buijtenen, J. & Price, R. Furfural—a promising platform for lignocellulosic biofuels. *ChemSusChem* **5**, 150–166 (2012).
- Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C. & Lorimer, G. H. Mechanism of Rubisco: the carbamate as general base. *Chem. Rev.* **98**, 549–562 (1998).
- Mani, K. Electrodialysis water splitting technology. *J. Membr. Sci.* **58**, 117–138 (1991).
- Davis, J. R., Chen, Y., Baygents, J. C. & Farrell, J. Production of acids and bases for ion exchange regeneration from dilute salt solutions using bipolar membrane electrodialysis. *ACS Sustainable Chem. Eng.* **3**, 2337–2342 (2015).
- Barve, P. P., Kamble, S. P., Joshi, J. B., Gupte, M. Y. & Kulkarni, B. D. Preparation of pure methyl esters from corresponding alkali metal salts of carboxylic acids using carbon dioxide and methanol. *Ind. Eng. Chem. Res.* **51**, 1498–1505 (2012).
- Dingyi, Y. & Yugen, Z. The direct carboxylation of terminal alkynes with carbon dioxide. *Green Chem.* **13**, 1275–1279 (2011).
- Kudo, K., Ikoma, F., Mori, S. & Sugita, N. Synthesis of glutamic acid salt from cesium 3-butenoate with carbon dioxide. *J. Jpn. Petrol. Inst.* **38**, 48–51 (1995).
- Vechorkin, O., Hirt, N. & Hu, X. Carbon dioxide as the C1 source for direct C–H functionalization of aromatic heterocycles. *Org. Lett.* **12**, 3567–3569 (2010).
- van Putten, R. J. et al. Hydroxymethylfurfural, a versatile platform chemical made from renewable resources. *Chem. Rev.* **113**, 1499–1597 (2013).
- Román-Leshkov, Y., Chheda, J. N. & Dumesic, J. A. Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science* **312**, 1933–1937 (2006).
- Binder, J. B. & Raines, R. T. Simple chemical transformation of lignocellulosic biomass into furans for fuels and chemicals. *J. Am. Chem. Soc.* **131**, 1979–1985 (2009).
- da Costa Lopes, A. M. & Bogel-Lukasik, R. Acidic ionic liquids as sustainable approach of cellulose and lignocellulosic biomass conversion without additional catalysts. *ChemSusChem* **8**, 947–965 (2015).
- Luterbacher, J. S. et al. Nonenzymatic sugar production from biomass using biomass-derived gamma-valerolactone. *Science* **343**, 277–280 (2014).
- Sheldon, R. A. Green and sustainable manufacture of chemicals from biomass: state of the art. *Green Chem.* **16**, 950–963 (2014).
- Hoydonckx, H., Van Rhijn, W., Van Rhijn, W., De Vos, D. & Jacobs, P. Furfural and derivatives. *Ullmann's Encyclopedia of Industrial Chemistry* **16**, 285–313 http://onlinelibrary.wiley.com/doi/10.1002/14356007.a12_119.pub2/full (Wiley-VCH, 2007).
- Taarning, E., Nielsen, I. S., Egeblad, K., Madsen, R. & Christensen, C. H. Chemicals from renewables: aerobic oxidation of furfural and hydroxymethylfurfural over gold catalysts. *ChemSusChem* **1**, 75–78 (2008).
- Thiyagarajan, S., Pukin, A., van Haveren, J., Lutz, M. & van Es, D. S. Concurrent formation of furan-2-, 5- and furan-2, 4-dicarboxylic acid: unexpected aspects of the Henkel reaction. *RSC Adv.* **3**, 15678–15686 (2013).
- Fischer, R. & Fišerová, M. One-step synthesis of furan-2, 5-dicarboxylic acid from furan-2-carboxylic acid using carbon dioxide. *ARKIVOC Online J. Org. Chem.* **4**, 405–412 (2013).
- Fraser, R. R., Mansour, T. S. & Savard, S. Acidity measurements in THF. V. Heteroaromatic compounds containing 5-membered rings. *Can. J. Chem.* **63**, 3505–3509 (1985).
- Renaud, P. & Fox, M. A. An electrochemical characterization of dianions: dilithiated carboxylic acids. *J. Am. Chem. Soc.* **110**, 5705–5709 (1988).
- Kudo, K. et al. Carboxylation of cesium 2-naphthoate in the alkali metal molten salts of carbonate and formate with CO₂ under high pressure. *J. Jpn. Petrol. Inst.* **38**, 40–47 (1995).
- Dibari, N., Aguayo, A. T., Bilbao, J., Urtiaga, A. & Ortiz, I. Membrane reactors for *in situ* water removal: a review of applications. *Ind. Eng. Chem. Res.* **52**, 10342–10354 (2013).

Acknowledgements We thank Stanford University and the Henry and Camille Dreyfus Foundation for support of this work through a Teacher-Scholar Award to M.W.K. G.R.D. gratefully acknowledges a fellowship through the Stanford Center for Molecular Analysis and Design, and T.Y. acknowledges a Postdoctoral Fellowship for Research Abroad through the Japan Society for the Promotion of Science. We thank T. Veltman for installation of the Parr reactor, S. Lynch for assistance with ²H NMR, and J. Du Bois for discussions. High-resolution mass spectrometry was performed at the Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University Mass Spectrometry.

Author Contributions M.W.K. and A.B. conceived the project. A.B., G.R.D. and T.Y. performed the experiments. M.W.K., A.B. and G.R.D. wrote the paper. All authors contributed to the analysis and interpretation of the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W.K. (mkanan@stanford.edu).

METHODS

Materials. Caesium carbonate (99.995%, trace metal basis), 2-thiophenecarboxylic acid (99%), benzoic acid-(phenyl-¹³C₆) (99 at% ¹³C), and tetrabutylammonium bromide (TBABr, 99%) were purchased from Sigma Aldrich; caesium carbonate ($\geq 99\%$, for large-scale reactions) and furan-2,5-dicarboxylic acid (99.6%) were purchased from Chem Impex International; dimethyl furan-2,5-dicarboxylate (99%) was purchased from Astatech; benzoic-d₅ acid-(phenyl-d₅) (98%) was purchased from Cambridge Isotope Laboratories; 2-furoic acid (98%), benzoic acid (99%) and anhydrous methyl alcohol (99.8%) were purchased from Acros Organics; sodium L-(+)-tartrate dihydrate (99.7%), benzene (HPLC grade) was purchased from Alfa Aesar; carbon dioxide (99.99%) was purchased from Praxair. The methanol was further dried over 3 Å molecular sieves before using. Reagent-grade benzene was dried by passing through an activated alumina column on an Innovative Technology PureSolv solvent purification system. N,N-dimethylformamide (DMF, 99.8%) was purchased from Fisher Scientific and dried by passing through an activated molecular sieve column. All other chemicals were used as received without further purification.

Equipment. Experiments under flowing CO₂ were performed in a Thermo Scientific Lindberg/Blue M tube furnace. Experiments under pressurized CO₂ or N₂ were performed in a 300-ml high-temperature, high-pressure Parr reactor (model 4561-HT-FG-SS-115-VS-2000-4848) equipped with a glass liner.

Structural analysis. ¹H-NMR, ²H-NMR and ¹³C-NMR spectra were recorded at 23 °C on a Varian Unity Inova 600 MHz spectrometer, a Varian Unity Inova 500 MHz spectrometer, a Varian Direct Drive 400 MHz spectrometer, a Varian Mercury 400 MHz spectrometer, or a Varian Unity Inova 300 MHz spectrometer. ¹H chemical shifts (δ) are reported in parts per million downfield from tetramethylsilane and referenced to residual protium in the NMR solvent (D₂O, $\delta = 4.79$, CDCl₃, $\delta = 7.26$). ²H chemical shifts (δ) are reported in parts per million downfield from tetramethylsilane and referenced to deuterium in the NMR solvent (D₂O, $\delta = 4.71$). ¹³C chemical shifts (δ) are reported in parts per million downfield from tetramethylsilane and referenced to carbon resonances in the NMR solvent (CDCl₃, $\delta = 77.16$, centre line), to added methanol ($\delta = 49.00$), or to added tetramethylsilane ($\delta = 0.00$).

High-resolution mass spectra were recorded on a Waters Aquity UPLC and Thermo Exactive Orbitrap mass spectrometer by direct injection electrospray ionization–mass spectrometry (ESI-MS).

Preparation of reactant mixtures consisting of caesium carboxylate + 0.55 equivalents Cs₂CO₃. The carboxylic acid (2-furoic acid, thiophene-2-carboxylic acid or benzoic acid) and 1.05 equivalent of Cs₂CO₃ were dissolved in the minimum amount of deionized water and evaporated to dryness by heating at 150 °C for at least 2 h. The solid mixture was cooled and used directly for C–H carboxylation reactions. We note that residual moisture in the reactant mixture reduces the yield of the carboxylation reaction.

C–H carboxylation under flowing CO₂. The appropriate amount of reactant mixture was weighed out into a quartz boat and placed in the tube furnace. The furnace was heated to the desired temperature under CO₂ flowing at 40 ml min⁻¹ for a given period of time. The sample was cooled to ambient temperature, dissolved in D₂O and filtered through a 0.2 µm polytetrafluoroethylene (PTFE) syringe filter to prepare samples for NMR analysis. The product yields were calculated from integration of the ¹H NMR peaks using sodium L-(+)-tartrate dihydrate as an internal standard. Representative spectra and data are shown in Extended Data Figs 1 and 2 and Extended Data Table 1a.

C–H carboxylation under pressurized CO₂. The reactant mixture was charged into the Parr reactor equipped with a glass liner. The reactor was sealed and then evacuated and backfilled with CO₂ three times. It was then filled to a final CO₂ pressure at ambient temperature corresponding to a pressure of 8 bar at the desired reaction temperature. The reactor was heated to the desired temperature, maintained at that temperature for a given period of time, and then cooled to ambient temperature and depressurized. The solid product mixture was dissolved in D₂O and analysed using NMR as described above. Representative spectra and data are shown in Extended Data Figs 3 and 4 and Extended Data Table 1a.

100-mmol-scale synthesis of furan-2,5-dicarboxylic acid at 1 atm of CO₂. To a 1-litre round-bottomed flask we added the 2-furoic acid (100 mmol, 11.21 g, 1.0 equiv.) and Cs₂CO₃ (125 mmol, 40.73 g, 1.25 equiv.) followed by 100 ml of deionized H₂O. The addition of water results in an acid–base reaction that liberates CO₂. (Caution: this reaction is exothermic and effervescent.) Once the reaction was complete, the water was removed in vacuum on a rotary evaporator (rotovap) at 75 °C and at 100 mTorr and 230 °C. The resulting off-white solid was scraped and ground into a fine white powder. In a fume hood, a reactor was assembled consisting of a rotovap with P₂O₅ in the collection flask connected to a Schlenk line and a eutectic salt bath (48.7 mol% NaNO₃, 51.3 mol% KNO₃). The eutectic salt bath was heated to 260 °C, and the 1-litre round-bottomed flask containing

the caesium furan-2-carboxylate and Cs₂CO₃ was attached to the rotovap. The joint was taped with black electrical tape and fitted with a green Keck clip, and the entire apparatus was backfilled with CO₂ gas. A short piece of tubing was connected in such a way as to deliver a slow stream of cooling air to the taped joint to prevent melting. The reaction was then dipped into the eutectic salt bath and rotated slowly for 48 h with a gentle flow of CO₂ through the bubbler of the Schlenk line. Over the course of the reaction, the solid initially melts, then turns blackish-brown and solidifies. Once complete, the reaction was slowly cooled to room temperature and detached from the rotovap. Disodium tartrate dihydrate (10 mmol, 2.31 g, 0.1 equiv.) was added followed by 200 ml of deionized water. An aliquot of the resulting solution was evaporated in vacuum then dissolved in D₂O. A ¹H NMR was obtained with the following distribution of products: caesium furan-2,5-dicarboxylate (71%), caesium malonate (2%), and caesium acetate (11%) (NMR yields). A repeat of the experiment gave the following distribution: caesium furan-2,5-dicarboxylate (69%), caesium malonate (7%), and caesium acetate (8%).

The product was isolated from the reaction mixture using the following procedure. First, the reaction was filtered through a pad of celite to remove insoluble material. The resulting solution was then acidified below a pH of 2 with sulfuric acid (15 ml concentrated acid). (Caution: this reaction is exothermic and effervescent.) The product precipitated from the solution and was collected on a Büchner funnel. The product was then dissolved in 1 litre of methanol and decolorized with activated carbon. The activated carbon was removed by filtering the solution through a pad of celite. The resulting solution was concentrated in vacuum and then triturated with 500 ml of ethyl acetate. The product was collected on a Büchner funnel, washed with ethyl acetate, and then dried in vacuum to afford a white crystalline powder (10.35 g, 66%).

C–H carboxylation of furan-2-carboxylate using K₂CO₃. 2-furoic acid (56 mg, 0.50 mmol, 1.0 equiv.), potassium isobutyrate (63 mg, 0.50 mmol, 1.0 equiv.), and K₂CO₃ (73 mg, 0.53 mmol, 1.05 equiv.) were dissolved in the minimum of water in a quartz boat, and evaporated to dryness by heating at 150 °C for 2 h under a stream of N₂. The sample was heated to 320 °C under CO₂ flowing at 40 ml min⁻¹ in the tube furnace for 8 h. The solid product mixture was dissolved in D₂O and analysed using ¹H NMR as described above. The yield of potassium FDCA²⁻ was 62%, and the conversion of furan-2-carboxylate was 76%. A representative spectrum is shown in Extended Data Fig. 5a.

A carboxylation reaction was also performed with 2-furoic acid (56 mg, 0.50 mmol, 1.0 equiv.), potassium acetate (49 mg, 0.50 mmol, 1.0 equiv.), and K₂CO₃ (73 mg, 0.53 mmol, 1.05 equiv.). The sample was heated to 300 °C under CO₂ flowing at 40 ml min⁻¹ in the tube furnace for 8 h. The yield of potassium FDCA²⁻ was 57%, and the conversion of furan-2-carboxylate was 96%.

Attempted carboxylation of caesium furan-2-carboxylate in DMF. A two-necked 25-ml round-bottomed flask was equipped with a reflux condenser, gas adaptor, PTFE-coated stir-bar, and a septum. The flask was charged with 2-furoic acid (115 mg, 1.03 mmol, 1.03 equiv.), Cs₂CO₃ (555 mg, 1.70 mmol, 1.70 equiv.), and water (2 ml). Once the resulting acid–base reaction subsided, the reaction vessel was heated to 150 °C in an oil bath under a stream of N₂. After 30 min the solution had dried out, and the reactor was then placed under vacuum and heated to 175 °C. At the same time, a heat gun was used to dry the rest of the apparatus. The reactor was then cooled to 125 °C and back-filled three times with CO₂. Using a syringe, dry DMF (2 ml) was added to the reaction. The reaction was then stirred for 12 h at 125 °C under a CO₂ atmosphere (1 bar). After 12 h, the reaction was concentrated under vacuum. NMR analysis of the residue in D₂O indicated no conversion of the starting material.

C–H carboxylation of benzene. Into a 20 ml vial we added a mixture of caesium carbonate (489 mg, 1.5 mmol, 1.0 equiv.) and caesium isobutyrate (220 mg, 1.0 mmol, 0.67 equiv.). This vial was placed into the Parr reactor, which was sealed and then evacuated and backfilled with CO₂ three times. Anhydrous benzene was injected into the reactor in an amount ranging from 10 ml to 35 ml depending on the desired final partial pressure of benzene. The reactor was then pressurized with CO₂ and heated to the desired temperature. The partial pressure of CO₂ at the final temperature was calculated assuming ideal behaviour. The partial pressure of benzene was calculated by subtracting the CO₂ pressure from the measured reactor pressure. After the indicated period of time (Extended Data Table 1b), the reactor was cooled to ambient temperature and depressurized. The crude product was dissolved in D₂O and analysed using NMR as described above. In addition to benzene carboxylation products, isobutyrate decomposition products were observed, which included formate, acetate, acetate carboxylation products (malonate and methane tricarboxylate), and insoluble char. Control experiments were performed in the absence of either benzene or gaseous CO₂. In the absence of benzene, only a trace amount of formate was observed and the main product was insoluble char from isobutyrate decomposition. In the absence of gaseous CO₂, a small amount of

benzoate was observed, which is attributed to the formation of CO₂ *in situ* from the decomposition of HCO₃⁻ that is formed by deprotonation of benzene with CO₃²⁻. Representative spectra and data are shown in Extended Data Fig. 5b and Extended Data Table 1b.

For comparison, an experiment was performed in the absence of caesium isobutyrate. A 20 ml vial with caesium carbonate (326 mg, 1.0 mmol) was placed into the Parr reactor, which was sealed and then evacuated and backfilled with CO₂ three times. 35 ml of anhydrous benzene was injected into the reactor. The reactor was pressurized with 15 bar of CO₂ and heated to 340 °C or 380 °C for 3 h and 8 h respectively. The reactor was then cooled to ambient temperature and depressurized. The crude product was dissolved in D₂O and analysed using NMR as described above. No reaction was observed in this case.

Thermal annealing of caesium furan-2-carboxylate and CO₂ in the absence of Cs₂CO₃. The Parr reactor was equipped with an oven-dried glass liner and charged with caesium furan-2-carboxylate (244 mg, 1.0 mmol). The reactor was sealed and then evacuated and backfilled with CO₂ three times. It was pressurized to 5 bar CO₂ and then heated to 200 °C, at which point the CO₂ pressure was 8 bar. After 2 h, the reactor was cooled to room temperature, vented, and disassembled. The residue was dissolved in D₂O and analysed using NMR. No FDCA²⁻ was formed in the reaction.

Recovery of caesium by acidic precipitation at the 10-mmol scale. The general procedure outlined previously (see sections ‘Preparation of reactant mixtures consisting of caesium carboxylate + 0.55 equivalents Cs₂CO₃,’ and ‘C–H carboxylation under pressurised CO₂’) was followed using 2-furoic acid (10 mmol, 1.12 g, 1.00 equiv.) and Cs₂CO₃ (10.5 mmol, 3.42 g, 1.05 equiv.). Once the reaction had completed and cooled to room temperature, the resulting solid was treated with 7 ml of 3 N HCl. The FDCA immediately precipitated from the reaction mixture as an off-white solid. The suspension was filtered through a glass frit (medium porosity) and washed with a minimum of deionized water (3 × 0.5 ml). The filtrate was then transferred to a 100-ml round-bottomed flask and the filter cake was transferred, washing with methanol, to a separate, tared, 100-ml round-bottomed flask. The flask containing the filter cake was evaporated to dryness to afford a yellow solid. The solid was analysed using ¹H NMR in acetone-d₆. NMR analysis indicated a crude isolated yield of 81% for FDCA, along with 8% residual unreacted 2-furoic acid. The flask containing the filtrate was evaporated to dryness to afford 3.74 g yellow solid. The solid was analysed using ¹H NMR in D₂O with an internal standard to quantify organic contaminants. 2-furoic acid and FDCA were present in an amount corresponding to <0.4% of the mass of the solid. Based on this analysis, the caesium was recovered in >99% yield as the CsCl salt.

Recovery of caesium by acidic precipitation at the 100-mmol scale. The general procedure outlined previously (see section ‘100-mmol-scale synthesis of furan-2,5-dicarboxylic acid at 1 atm of CO₂’) was followed using the following quantities of 2-furoic acid (100 mmol, 11.23 g, 1.00 equiv.) and Cs₂CO₃ (105 mmol, 34.36 g, 1.05 equiv.). Once the reaction had completed and cooled to room temperature, the resulting solid was treated with 110 ml of 2 N HCl. The FDCA immediately precipitated from the reaction mixture as an off-white solid. This reaction was filtered through a glass frit (medium porosity) and washed with a minimum of deionized water (3 × 5 ml). The filtrate was then transferred to a tared 500-ml round-bottomed flask and the filter cake was transferred and washed with methanol into a separate, tared, 500-ml round-bottomed flask. The filter cake and filtrate were massed and analysed as described above, yielding the following results: 69% crude isolated yield for FDCA; >98% recovery of caesium as CsCl.

H/D isotope exchange between furan-2-carboxylate and acetate-d₃. 2-Furoic acid (112 mg, 1.0 mmol), acetic acid-d₄ (58 µl, 1.0 mmol) and Cs₂CO₃ (682 mg, 2.1 mmol) were dissolved in the minimum amount of deionized water and evaporated to form a dry powder that consisted of 1 mmol caesium furan-2-carboxylate, 1 mmol caesium acetate-d₃ and 1.1 mmol of Cs₂CO₃. This material was heated in the Parr reactor to 200 °C under 2 bar of N₂ for 1 h. After cooling to room temperature, the product mixture was dissolved in D₂O and analysed by NMR. The integration of the furan-2-carboxylate peaks in the ¹H NMR spectrum using an internal standard indicated that the H/D scrambling was ~60% complete with substantially more scrambling at the 5 position than at the 3 and 4 positions. The presence of D at all positions was evident in the ²H NMR spectrum, the peak splitting of the ¹³C NMR spectrum, and in the high-resolution mass spectrum (Fig. 2d and Extended Data Figs 6b, 7a and 7b).

For comparison, an experiment was performed in the absence of Cs₂CO₃. 2-Furoic acid (112 mg, 1.0 mmol), acetic acid-d₄ (58 µl, 1.0 mmol) and Cs₂CO₃ (312 mg, 0.96 mmol) were dissolved in minimum amount of deionized water and evaporated to form an oil that consisted of caesium furan-2-carboxylate and caesium acetate-d₃. A ¹H NMR spectrum of this mixture is shown in

Extended Data Fig. 6a. This material was heated in the Parr reactor to 200 °C under 2 bar of N₂ for 1 h. After cooling to room temperature, the material was dissolved in D₂O and analysed by NMR. Integration of the furan-2-carboxylate peaks in the ¹H NMR spectrum using an internal standard indicated that the H/D scrambling was ~15% complete with substantial scrambling at the 5 position and almost no scrambling at the 3 and 4 positions. The presence of D at the 5 position was evident in the ²H NMR spectrum, the peak splitting of the ¹³C NMR spectrum, and in the high-resolution mass spectrum (Fig. 2d and Extended Data Fig. 8).

The mass spectrometry sample was prepared by adding 6 N HCl dropwise to the NMR sample until the clear solution turned into a suspension. The water was removed under vacuum and the residue was suspended in 2.5 ml of methanol. The suspension was allowed to stand till the solid particles settled. An aliquot of the supernatant liquid was further diluted with methanol and analysed directly by mass spectrometry.

Benzoic acid isotope exchange experiment. Benzoic acid-(phenyl-¹³C₆) (7.8 mg, 60.9 µmol), benzoic-d₅ acid-(phenyl-d₅) (7.8 mg, 61.3 µmol) and Cs₂CO₃ (41.5 mg, 127.4 µmol) were dissolved in 1 ml of deionized water and evaporated to form a dry powder that consists of caesium benzoate-(phenyl-¹³C₆), caesium benzoate-(phenyl-d₅) and 0.55 equivalents of Cs₂CO₃. This material was heated in the Parr reactor to 320 °C under 2 bar of N₂ for 1 h. After cooling to room temperature, the material was dissolved in D₂O and analysed by NMR. The ¹H spectra of the reactant mixture and the product mixture are shown in Fig. 3.

A control experiment was performed to test whether Cs₂CO₃ is necessary for isotopic scrambling. Benzoic acid-(phenyl-¹³C₆) (7.8 mg, 60.9 µmol), benzoic-d₅ acid-(phenyl-d₅) (7.8 mg, 61.3 µmol) and Cs₂CO₃ (19.8 mg, 60.7 µmol) were dissolved in 1 ml H₂O and evaporated to form a dry powder that consists of the caesium benzoate salts. After heating to 320 °C under 2 bar N₂ for 1 h, no H/D exchange was observed by ¹H NMR (Extended Data Fig. 9).

Esterification of caesium FDCA²⁻. Caesium FDCA²⁻ (420 mg, 1.0 mmol) was charged into the Parr reactor equipped with an oven-dried glass liner. The reactor was sealed and then evacuated and backfilled with CO₂ three times. Anhydrous methanol (100 ml) was injected into the reactor. The reactor was then pressurized with either 28.5 bar or 15 bar CO₂ and heated to 200 °C or 180 °C. After 30 min, the reactor was cooled to ambient temperature, vented, and disassembled. The reaction mixture was transferred to a 250-ml round-bottomed flask and the methanol was removed under vacuum on a rotary evaporator at 45 °C. The residue was washed twice with 5 mL CHCl₃ to dissolve the DMFD. The combined CHCl₃ washes were evaporated to afford DMFD as a white powder. The material was dissolved in CDCl₃ and analysed by ¹H NMR with TBABr as an internal standard. The remaining residue that was not dissolved in the CHCl₃ washes was dissolved in CD₃OD and analysed by ¹H NMR using TBABr as an internal standard. This material consists of FDCA²⁻, MMFD, and a small amount of additional DMFD.

Hydrolysis of dimethyl furan-2,5-dicarboxylate (DMFD). DMFD (184 mg, 1.0 mmol, 1 equiv.) and Cs₂CO₃ (326 mg, 1.0 mmol, 1 equiv.) were charged into a Parr reactor equipped with an oven-dried glass liner. The reactor was sealed and then evacuated and backfilled with CO₂ three times. Anhydrous methanol (100 ml) was injected into the reactor. The reactor was pressurized with 28.5 bar CO₂ and heated to 200 °C. The total pressure at 200 °C was 105 bar and the calculated CO₂ pressure was 45 bar. After 30 min, the reactor was cooled down to ambient temperature then vented and disassembled. The reaction mixture was transferred to a 250-ml round-bottomed flask and the methanol was removed under vacuum on a rotary evaporator at 45 °C. The residue was processed and analysed by ¹H NMR as described above for the esterification of FDCA²⁻.

Recycling Cs₂CO₃. The Parr reactor was equipped with an oven-dried glass liner and charged with caesium furan-2-carboxylate (244 mg, 1.0 mmol, 1.0 equiv.) and Cs₂CO₃ (179 mg, 0.55 mmol, 0.55 equiv.). The reactor was sealed and then evacuated and backfilled with CO₂ three times. The reactor was pressurized to 5 bar CO₂ and then heated to 200 °C, at which point the CO₂ pressure was 8 bar. After 5 h, the reactor was cooled and vented then evacuated and backfilled with N₂. To remove the water by-product, 10 ml of dry methanol was injected into the reactor to dissolve the reaction mixture. The methanol was removed by heating the reactor to 150 °C under vacuum. Subsequently, N₂ gas was flowed over the reaction mixture for 8 h by keeping the gas release valve of the reactor open. The reactor was cooled to ambient temperature and 90 ml of dry methanol was injected into it. The vessel was pressurized with 28.5 bar CO₂ and heated to 200 °C. After 30 min, the reactor was cooled to ambient temperature, depressurized and opened. The reaction mixture was diluted with 65 ml deionized water and extracted with CHCl₃ (2 × 65 ml). The combined organic layers were dried over Na₂SO₄ and concentrated under vacuum to afford the DMFD as a white solid and the yield was determined from ¹H NMR using TBABr as an internal standard. The aqueous extract was concentrated under reduced pressure to approximately a 2 ml solution

and transferred to the same glass liner used in the first cycle. To the solution was added 2-furoic acid equivalent to the amount of DMFD produced in the first cycle. The liner was resealed inside the reactor and heated to 150 °C under an atmosphere of N₂. The reactor was subsequently evacuated and backfilled with N₂, kept under a stream of N₂ at 150 °C for 8 h, and then cooled to ambient temperature. The reaction mixture was subjected to a second cycle of carboxylation followed by esterification following the same procedures described above. At the completion of the cycle, the solution was diluted with 65 ml deionized water and extracted with CHCl₃ (2 × 65 ml). The combined organic layers were dried over Na₂SO₄ and concentrated under vacuum to afford the DMFD (Extended Data Fig. 10a). Evaporation of the aqueous extract yielded the crude mixture of Cs₂FDCA, caesium 5-(methoxycarbonyl)furan-2-carboxylate and Cs₂CO₃. The amount of unreacted Cs₂FDCA and caesium 5-(methoxycarbonyl)furan-2-carboxylate were quantified by ¹H NMR using sodium L-(+)-tartrate dihydrate as an internal standard (Extended Data Fig. 10b).

NMR spectra. The NMR peaks in the spectra for the carboxylation reactions were assigned to different products by comparison to spectra for pure caesium salts obtained independently from the pure carboxylic acids. The resonances for these compounds are provided below for reference.

Caesium furan-2,5-dicarboxylate:

¹H-NMR (400 MHz, D₂O) δ 6.95 (s, 2H)

¹³C-NMR (100 MHz, D₂O) δ 166.1, 150.4, 115.8

Caesium acetate:

¹H-NMR (400 MHz, D₂O) δ 1.87 (s, 3H)

¹³C-NMR (100 MHz, D₂O) δ 181.4, 23.6

Caesium malonate:

¹H-NMR (400 MHz, D₂O) δ 3.09 (s, 2H)

¹³C-NMR (100 MHz, D₂O) δ 177.29, 49.03

Caesium thiophene-2,5-dicarboxylate:

¹H-NMR (300 MHz, D₂O) δ 7.4 (s, 2H)

¹³C-NMR (75 MHz, D₂O) δ 169.5, 144.8, 131.0

Caesium benzoate:

¹H-NMR (500 MHz, D₂O) δ 7.80 (d, J = 5 Hz, 2H), 7.41 (t, J = 7.5 Hz, 1H), 7.36 (t, J = 7.5 Hz, 2H)

¹³C-NMR (125 MHz, D₂O) δ 176.0, 137.0, 131.9, 129.5, 129.0

Caesium phthalate:

¹H-NMR (500 MHz, D₂O) δ 7.45 (dd, J = 5.7, 3.3 Hz, 2H), 7.37 (dd, J = 5.7, 3.3 Hz, 2H)

¹³C-NMR (125 MHz, D₂O) δ 177.8, 138.5, 129.5, 127.8

Caesium isophthalate:

¹H-NMR (500 MHz, D₂O) δ 8.25 (s, 1H), 7.89 (d, J = 7.5 Hz, 2H), 7.39 (t, J = 7.7 Hz, 1H)

¹³C-NMR (125 MHz, D₂O) δ 175.5, 137.1, 132.1, 129.7, 129.0

Caesium terephthalate:

¹H-NMR (500 MHz, D₂O) δ 7.80 (s, 4H)

¹³C-NMR (125 MHz, D₂O) δ 175.5, 139.4, 129.3

Caesium benzene-1,2,3-tricarboxylate:

¹H-NMR (500 MHz, D₂O) δ 7.47 (d, J = 7.7 Hz, 2H), 7.30 (t, J = 7.6 Hz, 1H)

¹³C-NMR (125 MHz, D₂O) δ 177.3, 177.1, 138.3, 137.5, 128.7, 127.8

Caesium benzene-1,3,5-tricarboxylate:

¹H-NMR (500 MHz, D₂O) δ 8.40 (s, 3H)

¹³C-NMR (125 MHz, D₂O) δ 175.0, 137.5, 132.2

Caesium benzene-1,2,4-tricarboxylate:

¹H-NMR (500 MHz, D₂O) δ 7.92 (s, 1H), 7.84 (d, J = 8 Hz, 1H), 7.48 (d, J = 8 Hz, 1H)

¹³C-NMR (125 MHz, D₂O) δ 177.6, 177.1, 175.1, 141.2, 138.1, 137.1, 129.9,

128.3, 127.6

Caesium benzene-1,2,4,5-tetracarboxylate:

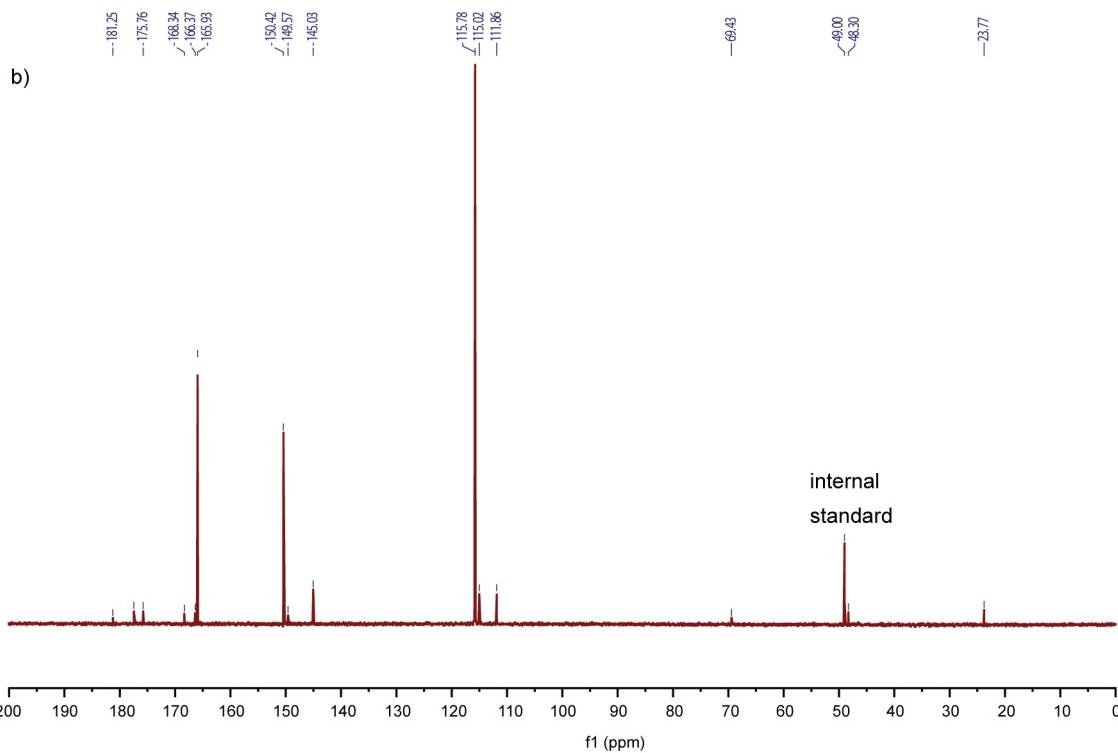
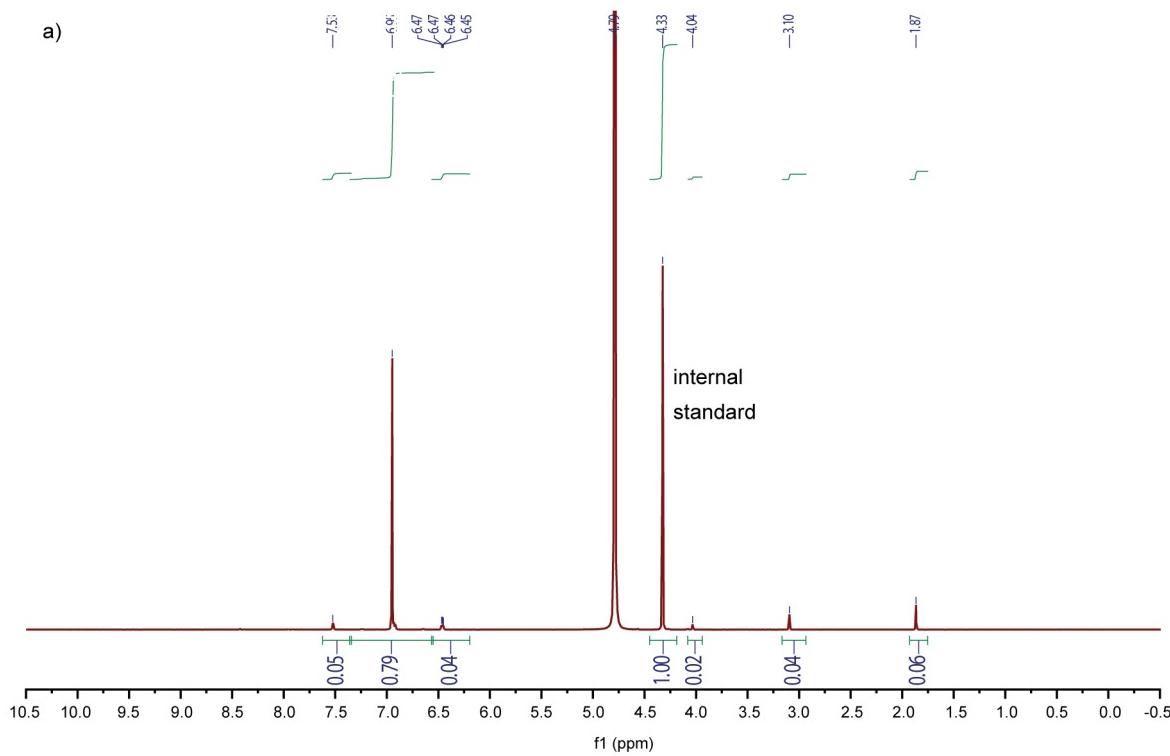
¹H-NMR (500 MHz, D₂O) δ 7.51 (s, 2H)

¹³C-NMR (125 MHz, D₂O) δ 176.8, 138.6, 126.6

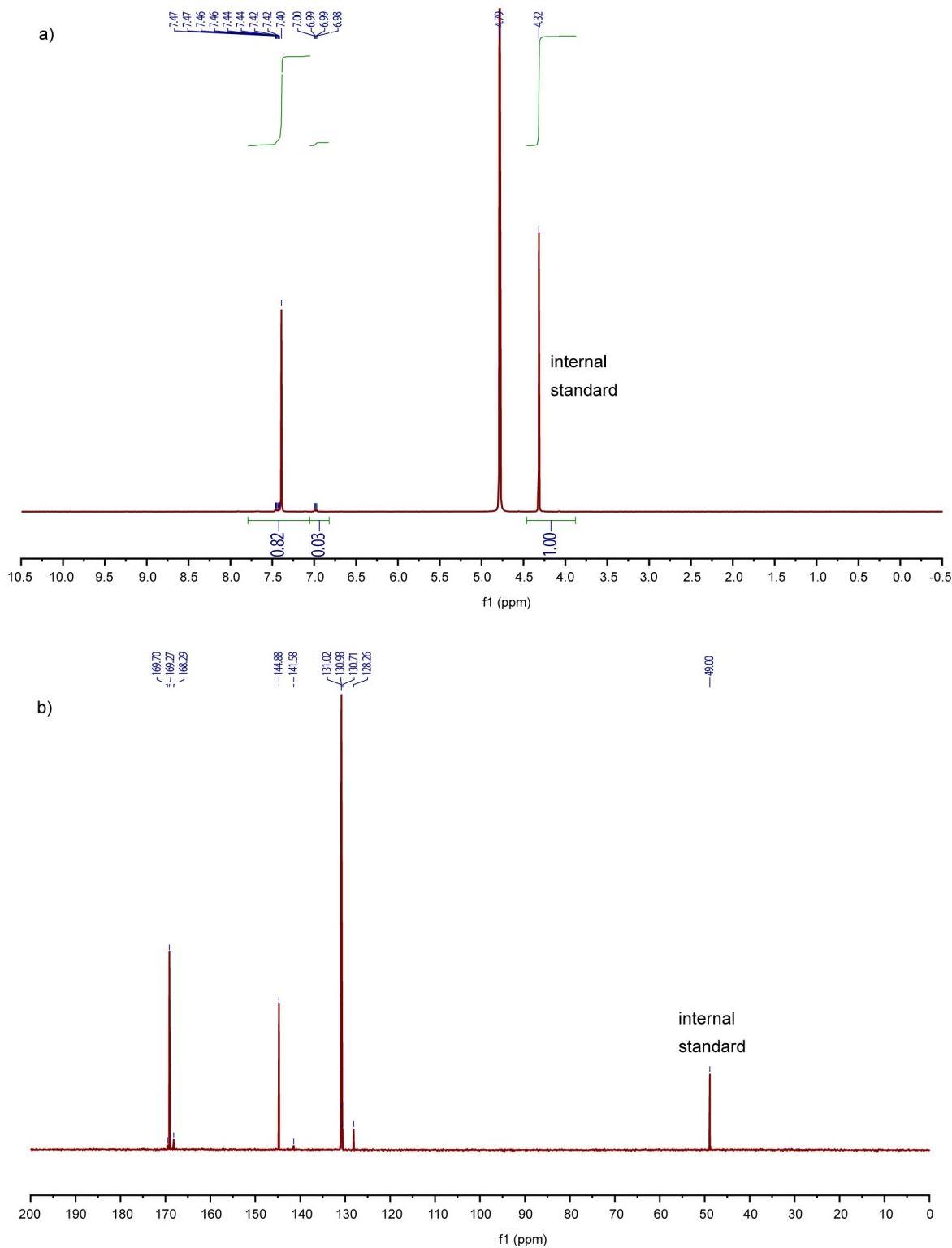
Dimethyl 2,5-furandicarboxylate:

¹H-NMR (300 MHz, CDCl₃) δ 7.20 (s, 2H), 3.91 (s, 6H)

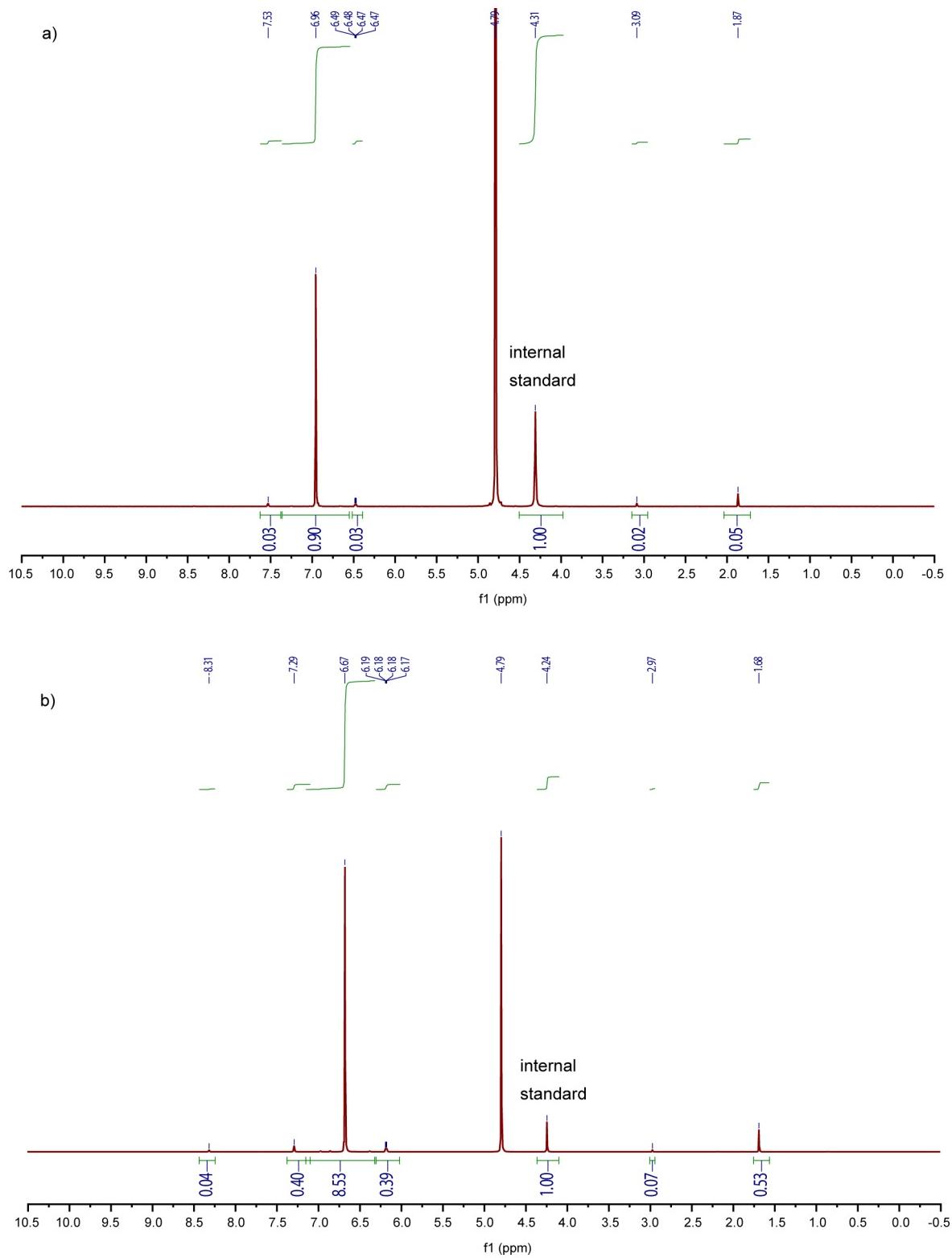
¹³C-NMR (75 MHz, CDCl₃) δ 158.5, 146.7, 118.6, 52.5



Extended Data Figure 1 | NMR spectra for the carboxylation of caesium furan-2-carboxylate under flowing CO₂. a, ¹H NMR (300 MHz) and b, ¹³C NMR (100 MHz) in D₂O of the crude product mixture after the reaction of 1 mmol caesium furan-2-carboxylate and 0.55 mmol Cs₂CO₃ under CO₂ flowing at 40 ml min⁻¹ at 260 °C for 12 h. f1 indicates the chemical shift, δ.

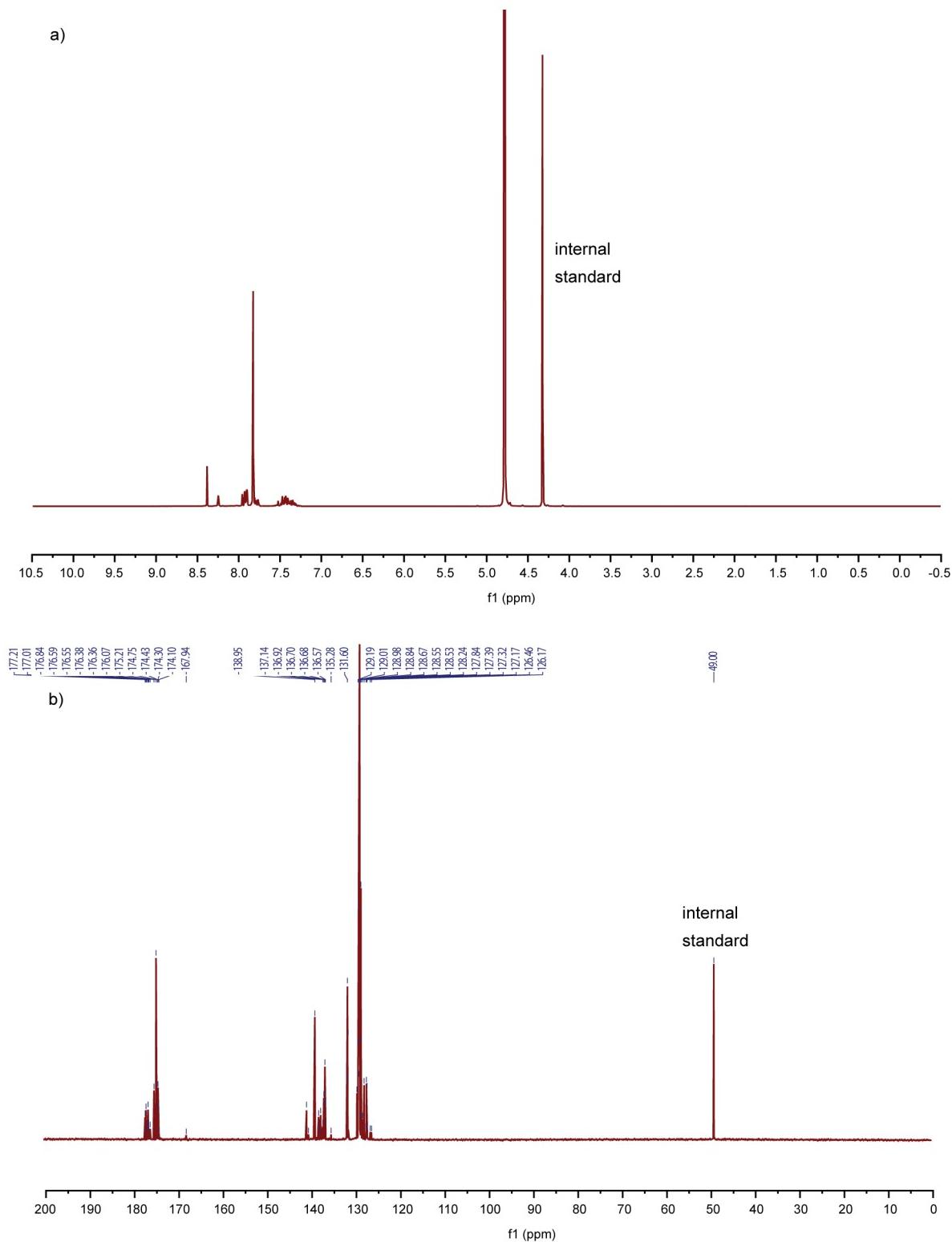


Extended Data Figure 2 | NMR spectra for the carboxylation of caesium thiophene-2-carboxylate. a, ^1H NMR (300 MHz) and b, ^{13}C NMR (100 MHz) in D_2O of the crude product mixture after the reaction of 1 mmol caesium thiophene-2-carboxylate and 0.55 mmol Cs_2CO_3 under CO_2 flowing at 40 ml min^{-1} at 325 °C for 12 h.

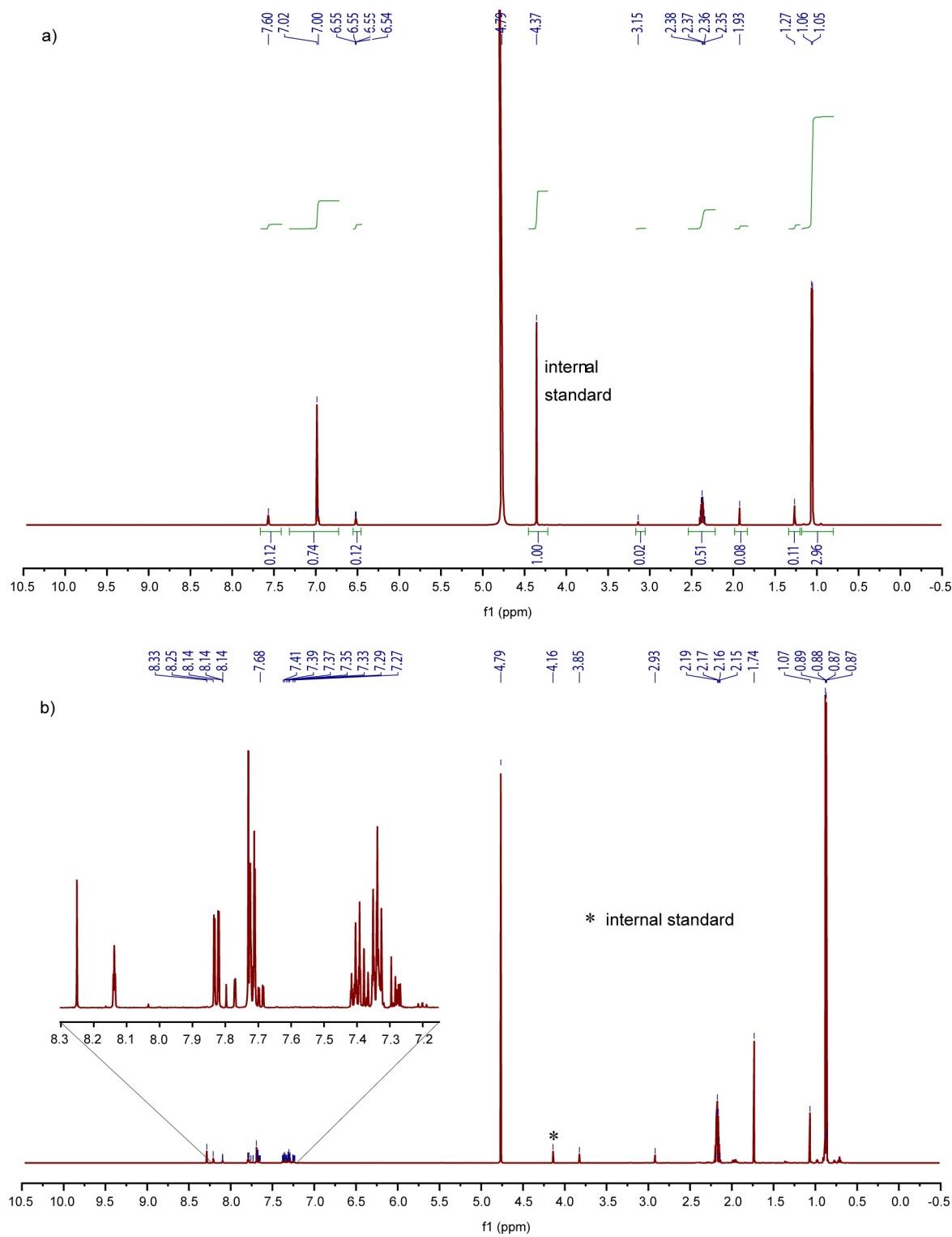


Extended Data Figure 3 | NMR spectra for the carboxylation of caesium furan-2-carboxylate in the Parr reactor. **a,** ^1H NMR (300 MHz) in D_2O of the crude product mixture after the reaction of 1 mmol caesium furan-2-carboxylate and 0.55 mmol Cs_2CO_3 under 8 bar CO_2 at 200 °C

for 5 h. **b,** ^1H NMR (300 MHz) in D_2O of the crude product mixture after the reaction of 10 mmol caesium furan-2-carboxylate and 5.5 mmol Cs_2CO_3 under 8 bar CO_2 at 200 °C for 10 h.

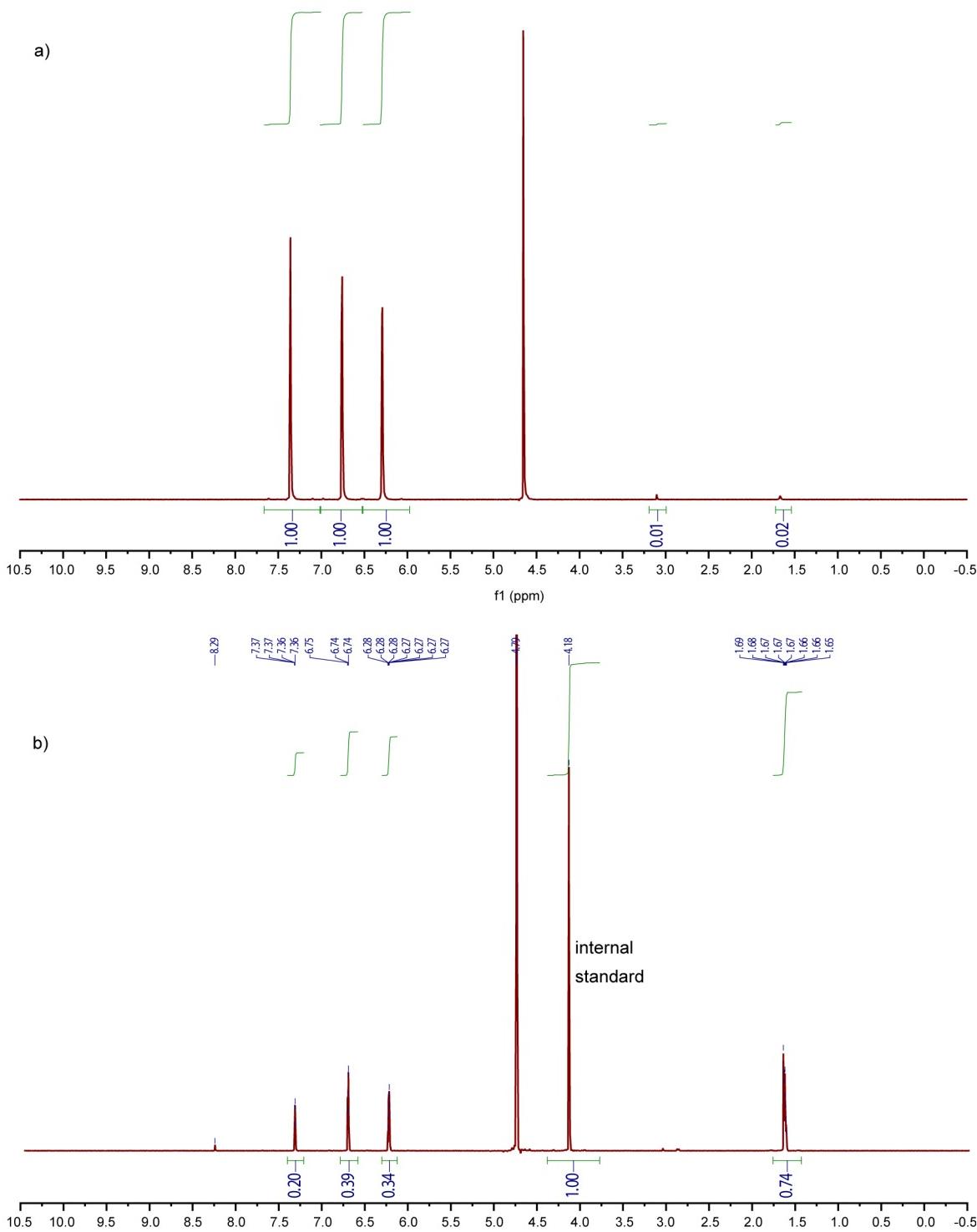


Extended Data Figure 4 | NMR spectra for the carboxylation of caesium benzoate. a, ^1H NMR (300 MHz) and b, ^{13}C NMR (100 MHz) in D_2O of the crude product mixture after the reaction of 1 mmol caesium benzoate and 0.55 mmol Cs_2CO_3 under 8 bar CO_2 at 320 °C for 5 h.

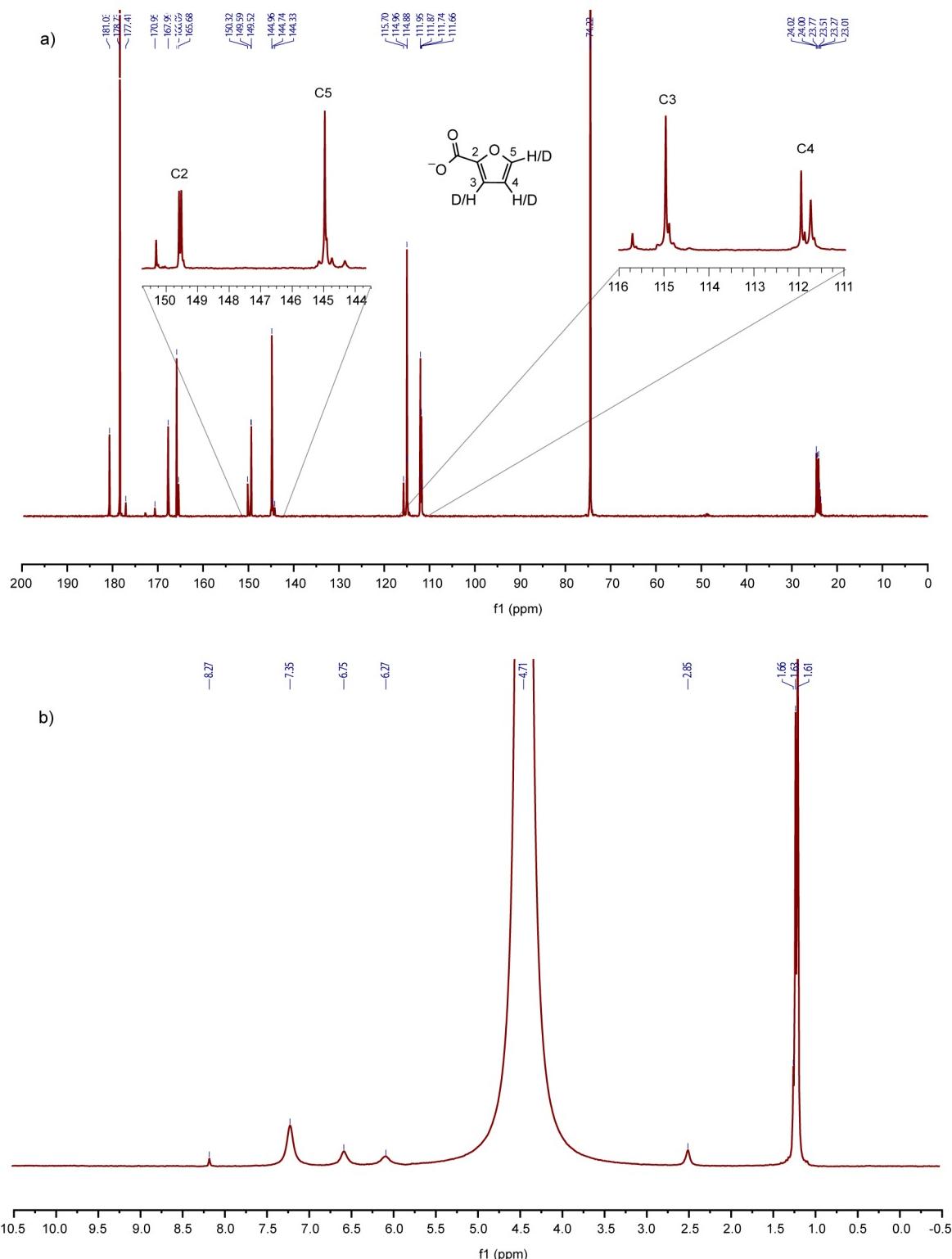


Extended Data Figure 5 | NMR spectra for the carboxylation of potassium furan-2-carboxylate and benzene. a, ^1H NMR (600 MHz) in D_2O of the crude product mixture after the reaction of 0.5 mmol potassium furan-2-carboxylate, 0.5 mmol potassium isobutyrate and

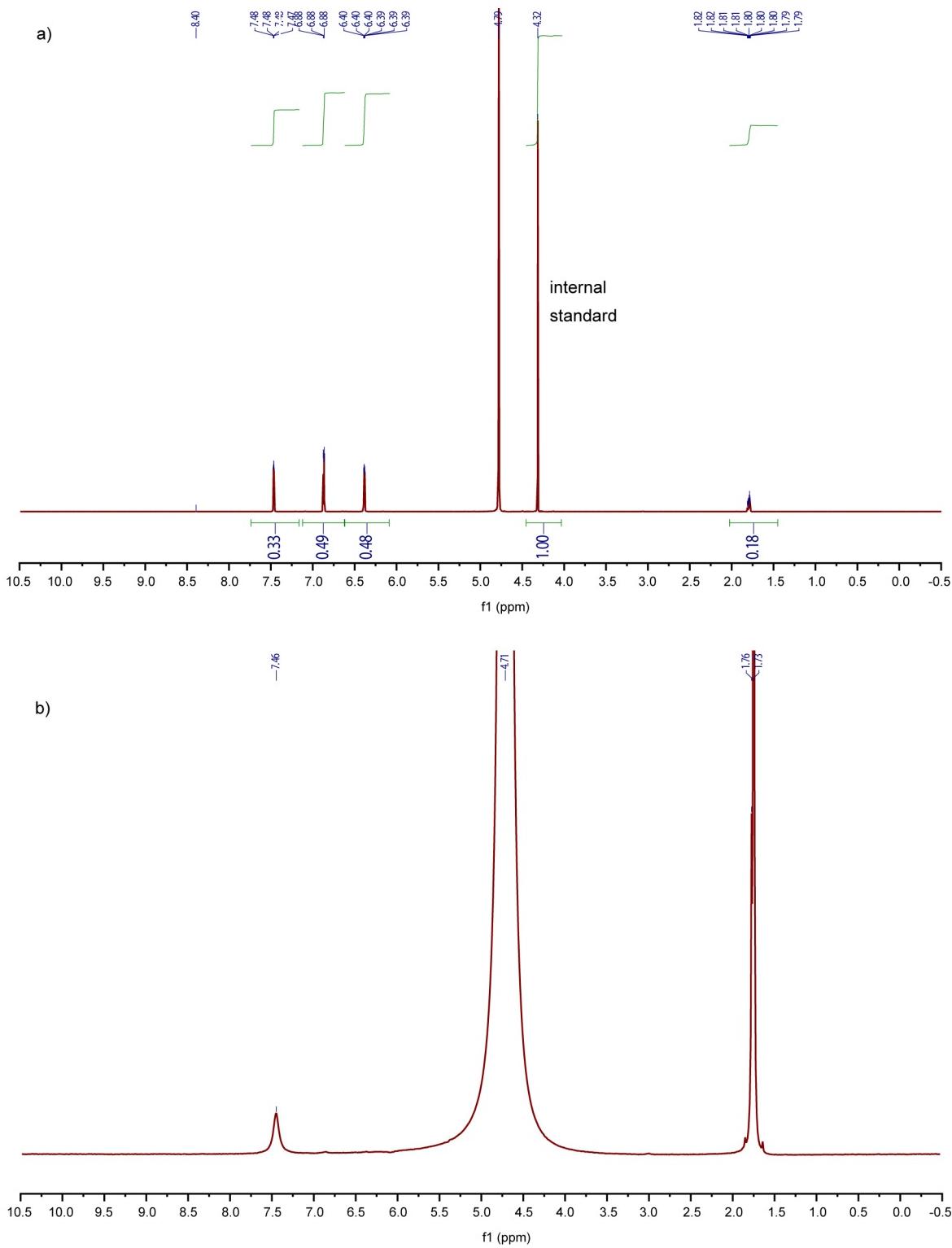
0.28 mmol K_2CO_3 under CO_2 flowing at 40 ml min^{-1} at 320°C for 8 h.
b, ^1H NMR (600 MHz) of the crude product mixture after the reaction in D_2O of a 1.5 mmol of caesium carbonate and 1 mmol caesium isobutyrate under 42 bar benzene and 31 bar CO_2 at 350°C for 8 h.



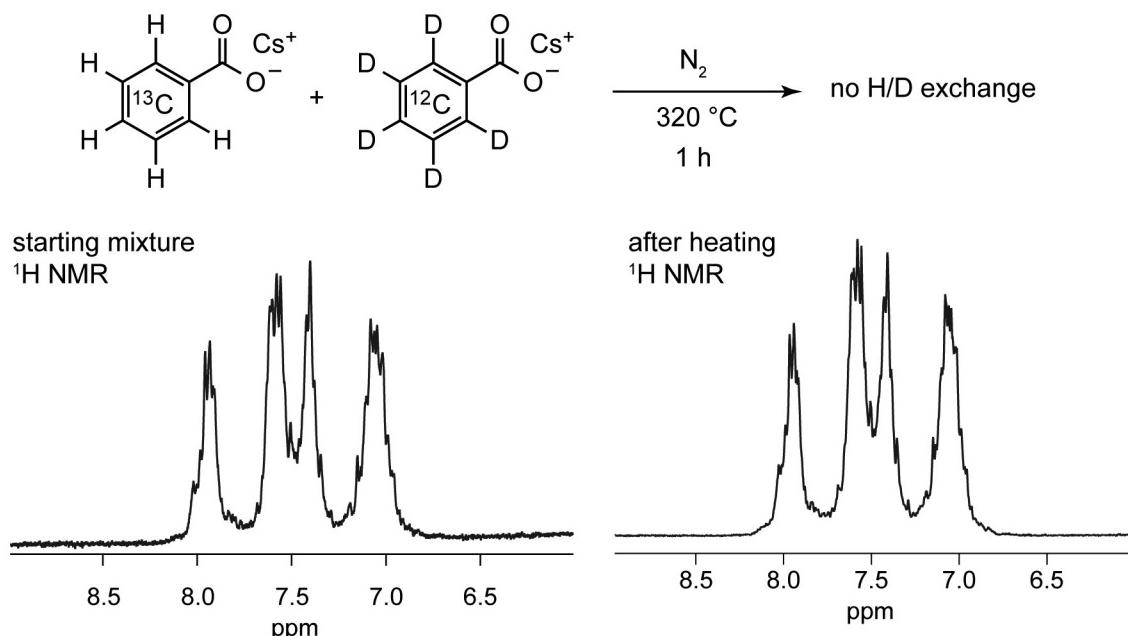
Extended Data Figure 6 | ^1H NMR spectra for H/D exchange between furan-2-carboxylate and deuterated acetate in the presence of Cs_2CO_3 . **a**, ^1H NMR (400 MHz) in D_2O of a 1:1 mixture of caesium furan-2-carboxylate and $\text{CD}_3\text{CO}_2\text{Cs}$. **b**, ^1H NMR (400 MHz) in D_2O of the crude product mixture after the reaction of a 1:1 mixture of caesium furan-2-carboxylate and $\text{CD}_3\text{CO}_2\text{Cs}$ with 0.55 equivalents Cs_2CO_3 at 200 °C under 2 bar N_2 for 1 h.



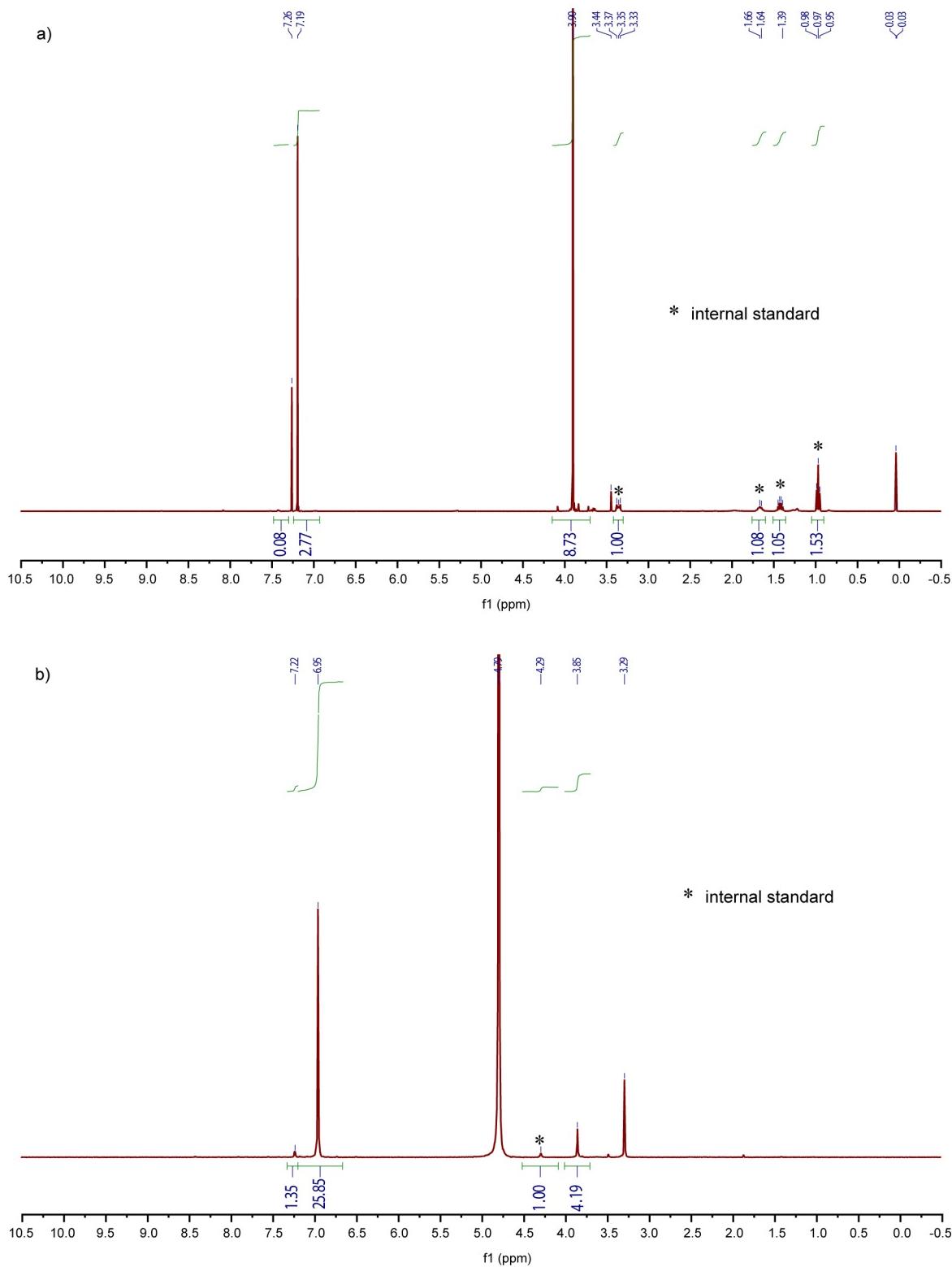
Extended Data Figure 7 | Additional NMR spectra for H/D exchange between furan-2-carboxylate and deuterated acetate in the presence of Cs_2CO_3 . a, ^{13}C NMR (75 MHz) and b, ^2H NMR (92 MHz) in D_2O of the crude product mixture after the reaction of a 1:1 mixture of caesium furan-2-carboxylate and $\text{CD}_3\text{CO}_2\text{Cs}$ with 0.55 equivalents Cs_2CO_3 at 200 °C under 2 bar N_2 for 1 h.



Extended Data Figure 8 | NMR spectra for H/D exchange between furan-2-carboxylate and deuterated acetate in the absence of Cs₂CO₃. a, ¹H NMR (400 MHz) and b, ²H NMR (92 MHz) in D₂O of the crude product mixture after the reaction of a 1:1 mixture of caesium furan-2-carboxylate and CD₃CO₂Cs at 200 °C under 2 bar N₂ for 1 h.



Extended Data Figure 9 | No H/D exchange is observed between differentially labelled caesium benzoates when heated to 320 °C in the absence of Cs₂CO₃.



Extended Data Figure 10 | NMR spectra for the Cs_2CO_3 recycling experiment. a, ^1H NMR (400 MHz) in CDCl_3 of the DMFD isolated after the second carboxylation/esterification sequence. b, ^1H NMR (400 MHz) in D_2O of the material recovered from the aqueous phase after the second carboxylation/esterification sequence.

Extended Data Table 1 | Additional C–H carboxylation data

Table 1a

entry	scale (mmol)	T	p(CO ₂)	time	FDCA ^{2–}	starting material	acetate	malonate	other
1	1	260 °C	flowing	6 h	57%	26%	4%	4%	9%
2	1	260 °C	flowing	12 h	76%	8%	4%	4%	8%
3	1	260 °C	flowing	20 h	76%	8%	5%	3%	8%
4	1	270 °C	flowing	4 h	66%	10%	7%	8%	9%
5	1	200 °C	8 bar	2 h	77%	18%	2%	1%	2%
6	1	200 °C	8 bar	5 h	89%	6%	3%	2%	--
7	1	200 °C	8 bar	7 h	89%	4%	2%	1%	4%
8	10	195 °C	8 bar	5 h	78%	11%	4%	1%	6%
9	10	195 °C	8 bar	10 h	81%	8%	4%	1%	6%
10	10	205 °C	8 bar	2 h	71%	7%	9%	3%	10%
11	10	215 °C	8 bar	2 h	69%	2%	14%	5%	10%
12	100	260 °C	1 bar	48 h	71%	3%	11%	2%	1%

Table 1b

entry	time	T	p(C ₆ H ₆)	p(CO ₂)	CO ₃ ^{2–} conv.	benzoate	phthalates	benzene tri + tetra carboxy- lates	acetate + carboxyla- tion pdts.	formate
1	10 h	340 °C	13 bar	31 bar	5%	2 µmol	14 µmol	14 µmol	205 µmol	17 µmol
2	9 h	340 °C	42 bar	31 bar	9%	42 µmol	29 µmol	11 µmol	238 µmol	24 µmol
3	2 h	350 °C	42 bar	31 bar	2%	19 µmol	3 µmol	1 µmol	27 µmol	9 µmol
4	8 h	350 °C	42 bar	31 bar	12%	46 µmol	41 µmol	17 µmol	148 µmol	28 µmol
5	8 h	360 °C	45 bar	32 bar	19%	39 µmol	69 µmol	35 µmol	306 µmol	42 µmol
6	1 h	380 °C	42 bar	33 bar	9%	30 µmol	36 µmol	13 µmol	192 µmol	146 µmol
7	8 h	350 °C	42 bar	—	0.6%	9 µmol	—	—	3 µmol	4 µmol

a, C–H carboxylation of caesium furan-2-carboxylate. b, C–H carboxylation of benzene. Formate, acetate, and acetate carboxylation products ('pdts') (malonate and methane tricarboxylate) arise from decomposition of isobutyrate.

Palladium-catalysed transannular C–H functionalization of alicyclic amines

Joseph J. Topczewski^{1*}, Pablo J. Cabrera^{1*}, Noam I. Saper¹ & Melanie S. Sanford¹

Discovering pharmaceutical candidates is a resource-intensive enterprise that frequently requires the parallel synthesis of hundreds or even thousands of molecules. C–H bonds are present in almost all pharmaceutical agents. Consequently, the development of selective, rapid and efficient methods for converting these bonds into new chemical entities has the potential to streamline pharmaceutical development^{1–4}. Saturated nitrogen-containing heterocycles (Alicyclic amines) feature prominently in pharmaceuticals, such as treatments for depression (paroxetine, amitifadine), diabetes (gliclazide), leukaemia (alvocidib), schizophrenia (risperidone, belaperidone), malaria (mefloquine) and nicotine addiction (cytisine, varenicline)⁵. However, existing methods for the C–H functionalization of saturated nitrogen heterocycles, particularly at sites remote to nitrogen, remain extremely limited^{6,7}. Here we report a transannular approach to selectively manipulate the C–H bonds of alicyclic amines at sites remote to nitrogen. Our reaction uses the boat conformation of the substrates to achieve palladium-catalysed amine-directed conversion of C–H bonds to C–C bonds on various alicyclic amine scaffolds. We demonstrate this approach by synthesizing new derivatives of several bioactive molecules, including varenicline.

Despite the ubiquity of alicyclic amines, there are very few methods available for the late-stage functionalization of these structures. Late-stage functionalization approaches are particularly valuable in the context of drug development, because they enable the rapid synthesis of analogues to optimize pharmacokinetic properties. Functionalization of C–H bonds using transition-metal catalysis offers a powerful approach for the late-stage functionalization of bioactive molecules^{1–4}, and recent progress in this field has led to thousands of new synthetic methods for selective C–H functionalization in a variety of molecular contexts^{2–4,8,9}. However, methods for the C–H functionalization of saturated nitrogen heterocycles remain extremely limited^{6,7}, and are dominated by functionalization of the highly activated C–H bonds α to nitrogen^{10–14} (Fig. 1b, left) or of C–H bonds on exocyclic alkyl groups^{15,16} (Fig. 1b, right). By contrast, here we describe an approach for achieving the C–H functionalization of alicyclic amine cores at sites remote from nitrogen (Fig. 1c) via nitrogen-directed transannular C–H activation.

We envisioned that coordination of the nitrogen of an alicyclic amine such as piperidine to palladium could enable selective transannular C–H activation^{17,18} to generate a bicyclo[2.2.1]palladacycle (exemplified by **1** in Fig. 2a). However, there are several challenges associated with this approach, including (1) the low equilibrium population of the

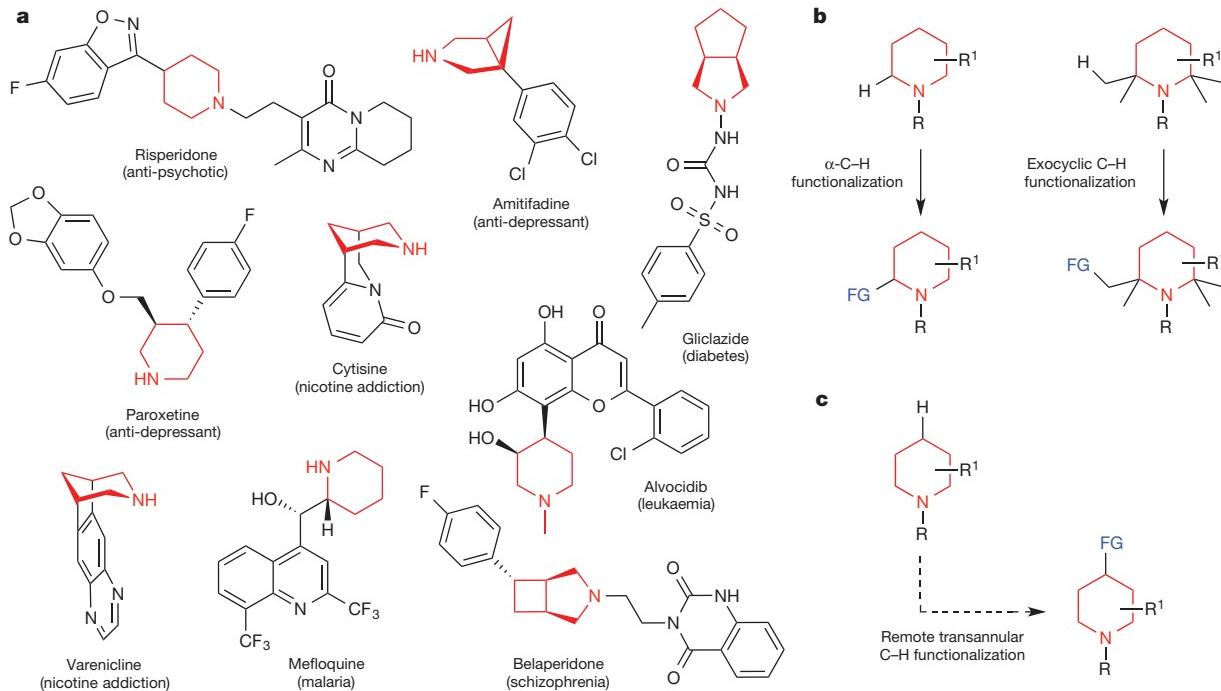


Figure 1 | Relevance of alicyclic amines and strategies for their late-stage functionalization. **a**, Representative pharmaceutical agents containing alicyclic amines. **b**, Previous synthetic approaches for the

late-stage functionalization of alicyclic amines. R, R¹, generic substituents; FG, functional group. **c**, Proposed approach for late-stage transannular C–H functionalization of alicyclic amines.

¹Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109, USA.

*These authors contributed equally to this work.

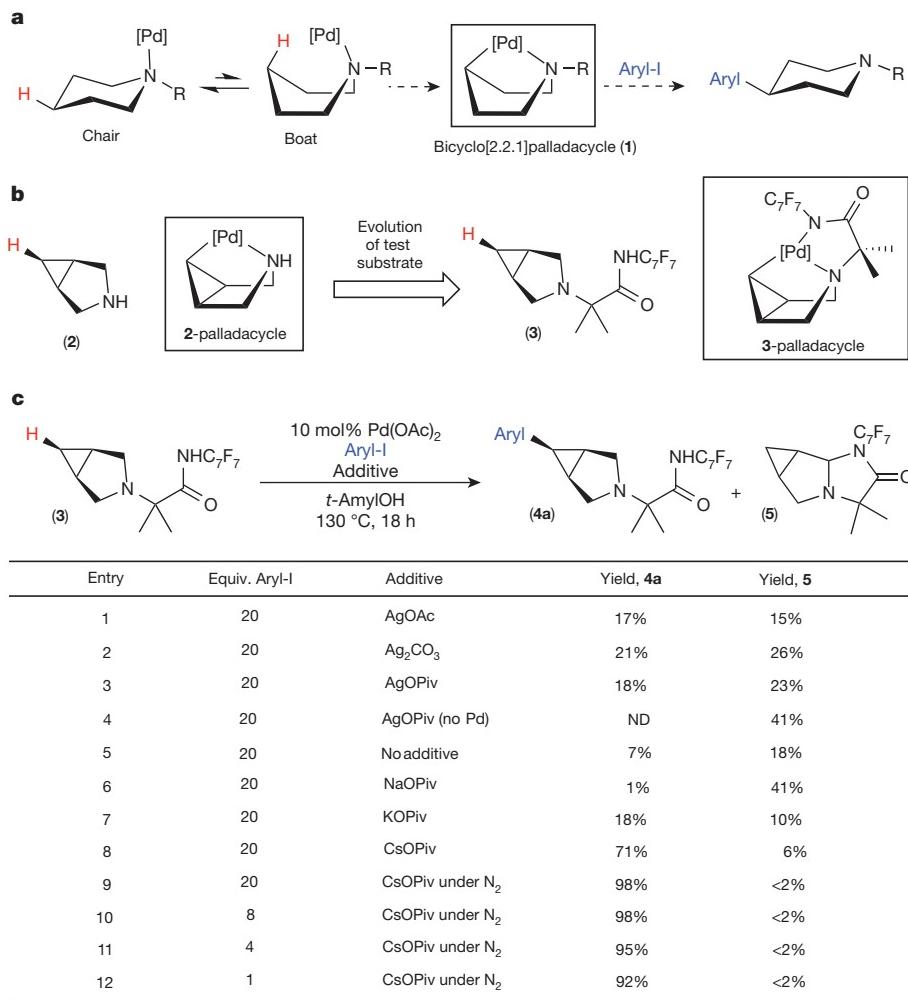


Figure 2 | Design and realization of transannular C–H activation of alicyclic amines. **a**, Conceptual approach for transannular C–H arylation of via a bicyclo[2.2.1]metallacycle intermediate. [Pd], Pd complex. **b**, Evolution of model substrate **2** to **3**. C₇F₇, p-CF₃C₆F₄.

c, Reaction optimization using 4-iodobiphenyl (Aryl-I). Boxed compounds are key conceptual intermediates. t-AmylOH, 2-methyl-2-butanol; ND, not detected. All yields determined by gas chromatography. See Supplementary Information for full details.

required boat conformer, (2) the requirement for cleavage of an unactivated secondary C(sp³)–H bond and (3) the potential susceptibility of the basic amine towards α-oxidation or N-oxidation. With these considerations in mind, we initially selected 3-azabicyclo[3.1.0]hexane (**2**) as a test substrate (Fig. 2b). We anticipated that the bicyclic core of **2** would prearrange it in a boat-like conformation and that the high s-character of the cyclopropyl C–H bonds should lower the barrier for C–H activation relative to a typical secondary C(sp³)–H site¹⁹.

The palladium (Pd)-catalysed reaction of **2** with 4-iodobiphenyl provided only traces of C–H arylated products under a variety of conditions. However, when a second coordinating group (an amide derived from a p-CF₃C₆F₄ aniline; Fig. 2c, **3**)^{20–22} was appended to nitrogen, the reaction afforded **4a** in modest to excellent yield. No products derived from C–H functionalization of the methyl groups of the fluoroamide directing group were observed in this reaction. This finding is in marked contrast to other reported applications of this directing group, in which C–H functionalization at β-methyl sites is strongly favoured^{21,22}, highlighting the complementarity of our approach that uses bidentate coordination of an sp³-hybridized nitrogen of an alicyclic amine substrate along with the fluoroamide to achieve selectivity (that is, transannular secondary C(sp³)–H functionalization).

The use of 10 mol% of Pd(OAc)₂ (Ac = acetate) and 1 equivalent of AgOAc (an additive commonly used to promote C–H arylation)²³ provided 17% of **4a** (Fig. 2c, entry 1). The modest yield of **4a** under

these conditions is due to competing formation of aminal **5**, which is thought to arise from α-oxidation of **3** to the corresponding iminium ion followed by intramolecular trapping with the amide nitrogen. The Ag additive mediates this transformation, and aminal **5** was obtained in 41% yield in the absence of Pd (Fig. 2c, entry 4). The role of the Ag carboxylate salt in these transformations is to regenerate the Pd carboxylate catalyst by abstraction of iodide from the Pd centre^{24,25}. As such, we hypothesized that the Ag salt could be replaced by a non-oxidizing metal carboxylate. A survey of alkali metal pivalate salts revealed that CsOPiv (Piv = pivalate) delivers the arylated product **4a** while suppressing the formation of aminal **5** (Fig. 2c, entry 8). Under the optimal conditions, **4a** was obtained in 92% yield as a single detectable stereoisomer (Fig. 2c, entry 12). X-ray crystallographic characterization of **4a** confirmed that the aryl group is installed on the concave face of the azabicycle (Fig. 3a).

This transannular C–H arylation reaction proceeds in high yield with aryl iodides bearing electron-donating, electron-neutral and electron-withdrawing substituents (products **4a–e**, Fig. 3a). Many traditionally sensitive functional groups are compatible with this system, including aryl bromides, unprotected phenols and aromatic aldehydes (products **4f–h**). Both electron-deficient and electron-rich nitrogen heterocycles can be installed (products **4i** and **4j**). Furthermore, a derivative of the amino acid phenylalanine can be coupled to the bicyclo[3.1.0] scaffold (product **4k**). Aryl bromides could also be used as the arylating reagent, albeit with reduced efficiency; for example, the

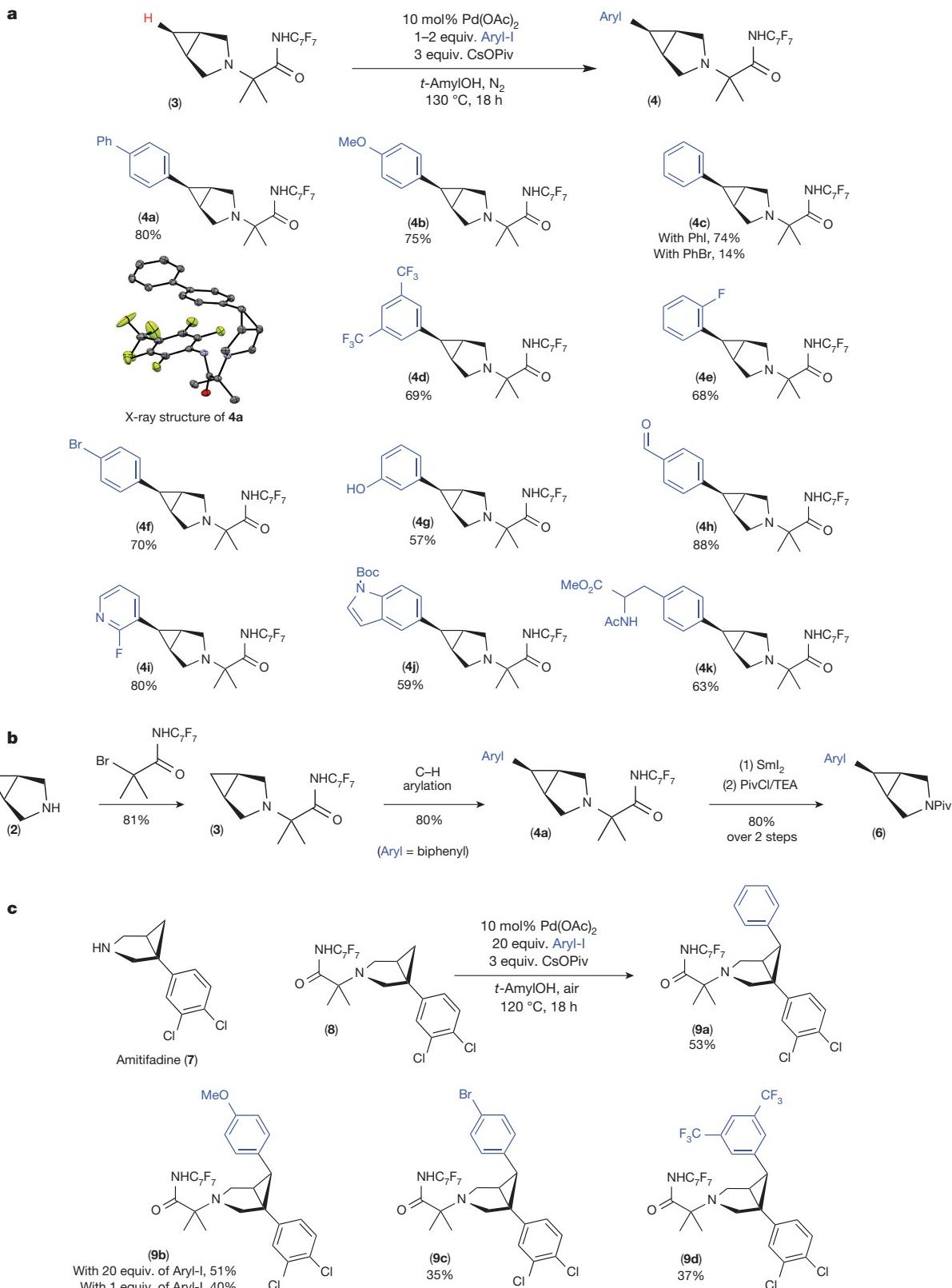


Figure 3 | Transannular C–H arylation of 3-azabicyclo[3.1.0]hexane core. **a**, Scope of C–H arylation with respect to the aryl iodide. Top, reaction studied; bottom, isolated products. Reaction to give **4c** with PhBr was conducted using 20 equiv. of PhBr ; yield determined by gas chromatography. Reaction to give **4j** was conducted under modified conditions. **b**, Relevant steps in overall transformation: installation of

directing group, C–H arylation and SmI_2 -mediated removal of directing group (Aryl = biphenyl). PivCl , pivaloyl chloride; TEA, triethylamine. **c**, C–H arylation applied to amitriptyline. Top left, structure of amitriptyline; top right, reaction studied; bottom, isolated products. All yields are reported for pure isolated material. See Supplementary Information for full details.

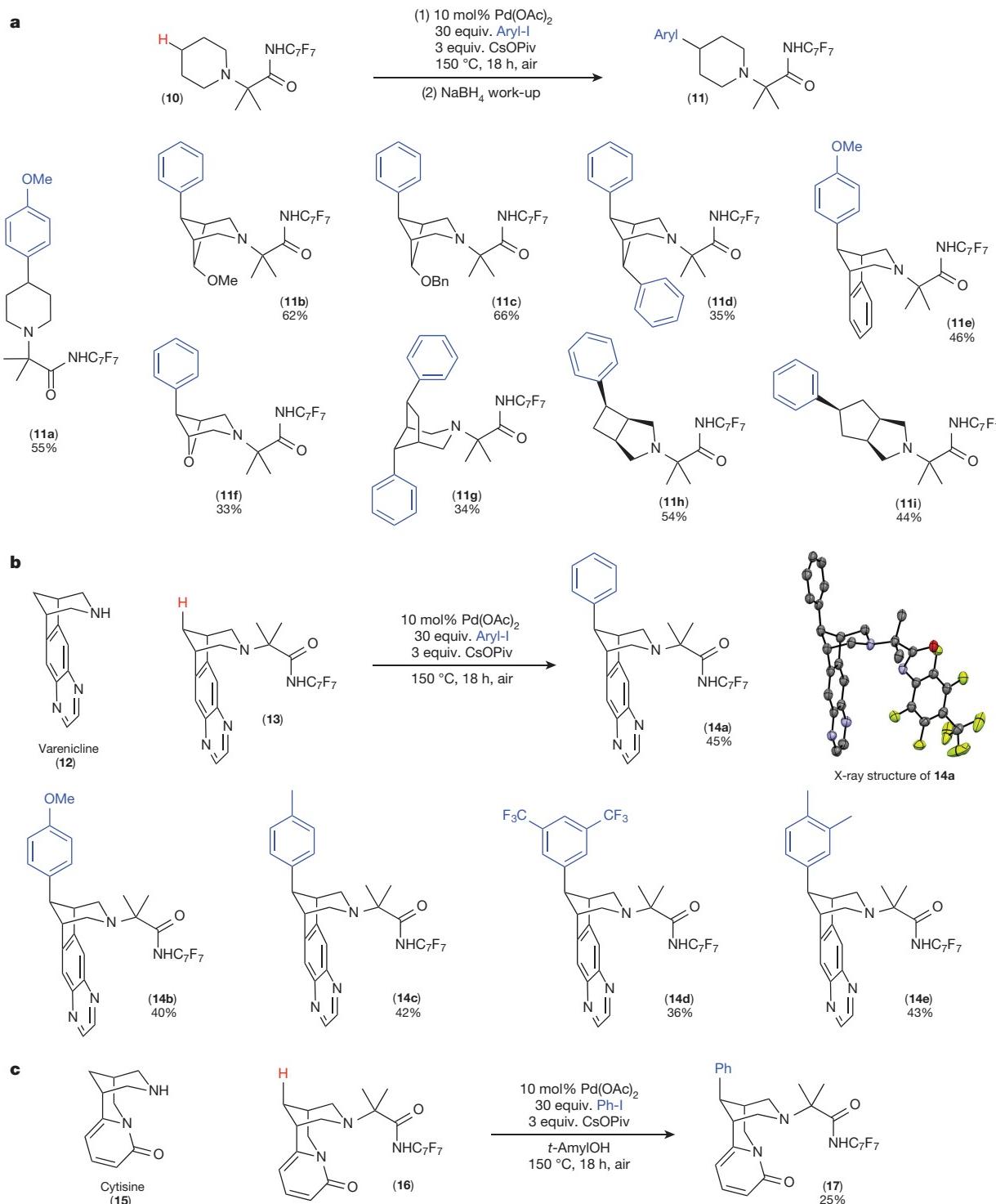


Figure 4 | Transannular C–H arylation of alicyclic amines. **a**, Scope of the C–H arylation reaction with respect to the amine. Top, reaction studied; bottom, isolated products. **b**, Application of this reaction to the derivatization of varenicline. Top left, structure of varenicline; top right, reaction studied; bottom, structure of varenicline derivatives. Reaction to

give **14d** was conducted under modified conditions. **c**, Application of this reaction to the derivatization of cytisine. Left, structure of cytisine (**15**); right, reaction studied. Yields are reported for pure isolated products. See Supplementary Information for full details.

use of phenyl bromide resulted in 14% yield of **4c** (see Supplementary Table 8 for full details).

The directing group can be removed in high yield via reductive cleavage with samarium diiodide (SmI_2). A 52% overall yield is obtained for the three relevant steps involved in converting **2** to **6** (81% for installation of the directing group, 80% for C–H arylation with 4-iodobiphenyl and 80% for removal of the directing group; Fig. 3b).

A particularly useful application of this method is in the late-stage derivatization of bioactive molecules. Selective C–H functionalization reactions on complex molecular scaffolds provide valuable opportunities for streamlining analogue generation and thereby accelerating structure–activity relationship studies^{4,8}. The bicyclo[3.1.0] scaffold appears in numerous pharmaceutical candidates, including the serotonin–noradrenaline–dopamine reuptake inhibitor amitriptyline (**7**)^{26,27}.

As shown in Fig. 3c, appending our directing group to **7** enables transannular C–H arylation to deliver the new amitifadine derivatives **9a–d**.

We next sought to expand this reaction from model substrate **3** to piperidine **10** (Fig. 4a). A thermodynamically unfavourable chair-boat isomerization of the piperidine ring in **10** is required before C–H activation (Fig. 2a) and is expected to add at least 6 kcal mol⁻¹ to the activation barrier relative to substrate **3** (ref. 28). Under the conditions optimized for **3**, the piperidine substrate **10** afforded only 12% yield of the C–H arylation product **11a**. However, increasing the temperature and removing the solvent led to a substantially improved 44% yield of **11a** (Fig. 4a). Amines derived from starting material **10** and product **11a** were formed as side products of this reaction (see Supplementary Fig. 2 for full details), but the reaction mixture could be cleanly converged to a mixture of starting material **10** and product **11a** via treatment with NaBH₄. Using this work-up procedure, product **11a** was isolated in 55% yield (Fig. 4a). Analogous conditions enabled the transannular C–H arylation of a variety of alicyclic amine derivatives, affording products of mono- and/or diarylation (**11b–i**; Fig. 4a). The structures of **11b–i** were established via a combination of NMR spectroscopy and X-ray crystallography.

Although the yields of **11b–i** are moderate in some cases, the *de novo* synthesis of many of these products would be challenging using traditional synthetic routes. The utility of this transformation is demonstrated in the late-stage C–H arylation of varenicline (**12**, Fig. 4b), a drug used to treat nicotine addiction. The fluoroamide group was appended to **12** to afford **13** in 81% yield. Under our standard C–H arylation conditions, **13** underwent transannular C–H arylation with a variety of aryl iodides to afford **14a–e**. The structure of **14a** was assigned by X-ray crystallography (Fig. 4b), which confirms that the aryl group is installed in an axial orientation. This point is particularly noteworthy because the synthesis of this stereoisomer would be challenging using other synthetic approaches²⁹. The C–H arylation of **13** with 4-iodo-*o*-xylene was conducted using 77 mg and 2.5 g of substrate, with nearly identical yields of **14e** (43% and 38% isolated yield, respectively). On the basis of the established synthesis of varenicline, an independent synthesis of these analogues by more traditional methods would require parallel multistep sequences³⁰. In a similar fashion, our method proved effective for the late-stage C–H functionalization of the natural product cytisine (**15**, a treatment for nicotine addiction), converting **16** to **17** in 25% yield (Fig. 4c). Again, the aryl group is selectively installed at the axial position in this transformation.

We have reported the transannular C–H arylation of a variety of alicyclic amines. The reaction exhibits high functional-group tolerance and enables the synthesis of new amino-acid derivatives (**4k**) as well as analogues of the pharmaceutical candidate amitifadine (**9a–d**), the drug varenicline (**14a–e**) and the natural product cytisine (**17**). We anticipate that a similar approach will ultimately be useful for the remote C–H functionalization of diverse cyclic and acyclic secondary amine scaffolds.

Received 3 August; accepted 21 December 2015.

Published online 17 February 2016.

- Wencel-Delord, J. & Glorius, F. C–H bond activation enables the rapid construction and late stage diversification of functional molecules. *Nature Chem.* **5**, 369–375 (2013).
- McMurray, L., O'Hara, F. & Gaunt, M. J. Recent developments in natural product synthesis using metal-catalysed C–H bond functionalization. *Chem. Soc. Rev.* **40**, 1885–1898 (2011).
- Yamaguchi, J., Yamaguchi, A. D. & Itami, K. C–H bond functionalization: emerging synthetic tools for natural products and pharmaceuticals. *Angew. Chem. Int. Ed.* **51**, 8960–9009 (2012).
- Godula, K. & Sames, D. C–H bond functionalization in complex organic synthesis. *Science* **312**, 67–72 (2006).
- Taylor, R. D., MacCoss, M. & Lawson, A. D. G. Rings in drugs. *J. Med. Chem.* **57**, 5845–5859 (2014).
- Asensio, G., Gonzalez-Nunez, M. E., Bernardini, C. B., Mello, R. & Adam, W. Regioselective oxyfunctionalization of unactivated tertiary and secondary C–H bonds of alkylamines by methyl(trifluoromethyl)dioxirane in acid medium. *J. Am. Chem. Soc.* **115**, 7250–7253 (1993).

- Affron, D. P., Davis, O. A. & Bull, J. A. Regio- and stereospecific synthesis of C–3 functionalized proline derivatives by palladium catalyzed directed C(sp³)–H arylation. *Org. Lett.* **16**, 4956–4959 (2014).
- Lyons, T. W. & Sanford, M. S. Palladium-catalyzed ligand-directed C–H functionalization reactions. *Chem. Rev.* **110**, 1147–1169 (2010).
- Chen, X., Engle, K. M., Wang, D. & Yu, J.-Q. Palladium(II)-catalyzed C–H activation/C–C cross-coupling reactions: versatility and practicality. *Angew. Chem. Int. Ed.* **48**, 5094–5115 (2009).
- Mitchell, E. A., Peschiali, A., Lefevre, N., Meerpoel, L. & Maes, B. U. W. Direct α -functionalization of saturated cyclic amine. *Chemistry* **18**, 10092–10142 (2012).
- Pastine, S. J., Gribkov, D. V. & Sames, D. sp³ C–H bond arylation directed by amidine protecting group: α -arylation of pyrrolidines and piperidines. *J. Am. Chem. Soc.* **128**, 14220–14221 (2006).
- He, J., Hamann, L. G., Davies, H. M. L. & Beckwith, R. E. J. Late-stage C–H functionalization of complex alkaloids and drug molecules via intermolecular rhodium–carbenoid insertion. *Nature Commun.* **6**, 5943 (2015).
- Shi, L. & Xia, W. Photoredox functionalization of C–H bonds adjacent to a nitrogen atom. *Chem. Soc. Rev.* **41**, 7687–7697 (2012).
- Spangler, J. E., Kobayashi, Y., Verma, P., Wang, D.-H. & Yu, J.-Q. α -Arylation of saturated azacycles and N-methylamines via palladium(II)-catalyzed C(sp³)–H coupling. *J. Am. Chem. Soc.* **137**, 11876–11879 (2015).
- McNally, A., Haffemayer, B., Collins, B. S. L. & Gaunt, M. J. Palladium-catalysed C–H activation of aliphatic amines to give strained nitrogen heterocycles. *Nature* **510**, 129–133 (2014).
- Lee, M. & Sanford, M. S. Platinum-catalyzed terminal-selective C(sp³)–H oxidation of aliphatic amines. *J. Am. Chem. Soc.* **137**, 12796–12799 (2015).
- Bercaw, J. E. et al. Robotic lepidoptery: structural characterization of (mostly) unexpected palladium complexes obtained from high-throughput catalyst screening. *Organometallics* **28**, 5017–5024 (2009).
- Cui, W. et al. Palladium-catalyzed remote C(sp³)–H arylation of 3-pinanamine. *Org. Lett.* **16**, 4288–4291 (2014).
- Giri, R., Chen, X. & Yu, J.-Q. Palladium-catalyzed asymmetric iodination of unactivated C–H bonds under mild conditions. *Angew. Chem. Int. Ed.* **44**, 2112–2115 (2005).
- Rouquet, G. & Chatani, N. Catalytic functionalization of C(sp²)–H and C(sp³)–H bonds by using bidentate directing groups. *Angew. Chem. Int. Ed.* **52**, 11726–11743 (2013).
- Wasa, M. et al. Ligand-enabled methylene C(sp³)–H bond activation with a Pd(II) catalyst. *J. Am. Chem. Soc.* **134**, 18570–18572 (2012).
- He, J. et al. Ligand-controlled C(sp³)–H arylation and olefination in synthesis of unnatural chiral α -amino acids. *Science* **343**, 1216–1220 (2014).
- Zaitsev, V. G., Shabashov, D. & Daugulis, O. Highly regioselective arylation of sp³ C–H bonds catalyzed by palladium acetate. *J. Am. Chem. Soc.* **127**, 13154–13155 (2005).
- Nadres, E. T., Santos, G. I. F., Shabashov, S. & Daugulis, O. Scope and limitations of auxiliary-assisted, palladium-catalyzed arylation and alkylation of sp² and sp³ C–H bonds. *J. Org. Chem.* **78**, 9689–9714 (2013).
- Lafrance, M. & Fagnou, K. Palladium-catalyzed benzene arylation: incorporation of catalytic pivalic acid as a proton shuttle and a key element in catalyst design. *J. Am. Chem. Soc.* **128**, 16496–16497 (2006).
- Beer, B. et al. DOV 216,303, a “triple” reuptake inhibitor: safety, tolerability, and pharmacokinetic profile. *J. Clin. Pharmacol.* **44**, 1360–1367 (2004).
- Epstein, J. W. et al. 1-Aryl-3-azabicyclo[3.1.0]hexanes, a new series of nonnarcotic analgesic agents. *J. Med. Chem.* **24**, 481–490 (1981).
- Juaristi, E. *Conformational Behavior of Six-Membered Rings: Analysis, Dynamics, and Stereoelectronic Effects* (Wiley–VCH, 1995).
- Hirsch, J. A. in *Topics in Stereochemistry* Vol. 1 (eds Allinger, N. L. & Eliel, E. L.) 199–222 (John Wiley & Sons, 1967).
- Singer, R. A., McKinley, J. D., Barbe, G. & Farlow, R. A. Preparation of 1,5-methano-2,3,4,5-tetrahydro-1H-3-benzazepine via Pd-catalyzed cyclization. *Org. Lett.* **6**, 2357–2360 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge J. W. Kampf for X-ray crystallographic analyses of **4a**, **11b**, an analogue of **11g**, **11h** and **14a**. J.T.T. was supported by an NIH post-doctoral fellowship (F32 GM109479). This work was supported by NIGMS grant GM073836. We acknowledge funding from NSF grant CHE-0840456 for X-ray instrumentation.

Author Contributions J.T.T., P.J.C. and N.I.S. discovered and developed the reaction. J.T.T., P.J.C. and M.S.S. conceived and designed the investigations. M.S.S. directed and supported the research. J.T.T., P.J.C. and M.S.S. wrote and revised the manuscript.

Author Information Metrical parameters for the structures are available free of charge from the Cambridge Crystallographic Data Centre under reference numbers CCDC-1401221, 1401222, 1440132, 1416579 and 1416516. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S.S. (mssanfor@umich.edu).

The terrestrial biosphere as a net source of greenhouse gases to the atmosphere

Hanqin Tian¹, Chaoqun Lu^{1,2}, Philippe Ciais³, Anna M. Michalak⁴, Josep G. Canadell⁵, Eri Saikawa⁶, Deborah N. Huntzinger⁷, Kevin R. Gurney⁸, Stephen Sitch⁹, Bowen Zhang¹, Jia Yang¹, Philippe Bousquet³, Lori Bruhwiler¹⁰, Guangsheng Chen¹¹, Edward Dlugokencky¹⁰, Pierre Friedlingstein¹², Jerry Melillo¹³, Shufen Pan¹, Benjamin Poulter¹⁴, Ronald Prinn¹⁵, Marielle Saunois³, Christopher R. Schwalm^{7,16} & Steven C. Wofsy¹⁷

The terrestrial biosphere can release or absorb the greenhouse gases, carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O), and therefore has an important role in regulating atmospheric composition and climate¹. Anthropogenic activities such as land-use change, agriculture and waste management have altered terrestrial biogenic greenhouse gas fluxes, and the resulting increases in methane and nitrous oxide emissions in particular can contribute to climate change^{2,3}. The terrestrial biogenic fluxes of individual greenhouse gases have been studied extensively^{4–6}, but the net biogenic greenhouse gas balance resulting from anthropogenic activities and its effect on the climate system remains uncertain. Here we use bottom-up (inventory, statistical extrapolation of local flux measurements, and process-based modelling) and top-down (atmospheric inversions) approaches to quantify the global net biogenic greenhouse gas balance between 1981 and 2010 resulting from anthropogenic activities and its effect on the climate system. We find that the cumulative warming capacity of concurrent biogenic methane and nitrous oxide emissions is a factor of about two larger than the cooling effect resulting from the global land carbon dioxide uptake from 2001 to 2010. This results in a net positive cumulative impact of the three greenhouse gases on the planetary energy budget, with a best estimate (in petagrams of CO_2 equivalent per year) of 3.9 ± 3.8 (top down) and 5.4 ± 4.8 (bottom up) based on the GWP100 metric (global warming potential on a 100-year time horizon). Our findings suggest that a reduction in agricultural methane and nitrous oxide emissions, particularly in Southern Asia, may help mitigate climate change.

The concentration of atmospheric CO_2 has increased by nearly 40% since the start of the industrial era, while CH_4 and N_2O concentrations have increased by 150% and 20%, respectively^{3,7,8}. Although thermo-genic sources (for example, fossil fuel combustion and usage, cement production, geological and industrial processes) represent the single largest perturbation of climate forcing, biogenic sources and sinks also account for a significant portion of the land-atmosphere exchange of these gases. Land biogenic greenhouse gas (GHG) fluxes are those originating from plants, animals and microbial communities, with changes driven by both natural and anthropogenic perturbations (see Methods). Although the biogenic fluxes of CO_2 , CH_4 and N_2O have been individually measured and simulated at various spatial and temporal scales, an overall GHG balance of the terrestrial biosphere is lacking³. But simultaneous quantification of the fluxes of these three gases is needed for developing effective climate change mitigation strategies^{9,10}.

In the analysis that follows, we use a dual-constraint approach from 28 bottom-up (BU) studies and 13 top-down (TD) atmospheric inversion studies to constrain biogenic fluxes of the three gases. We generate decadal mean estimates and s.d. of CO_2 , CH_4 and N_2O fluxes (mean \pm s.d., with s.d. being the square root of the quadratic sum of standard deviations reported by individual studies) in land biogenic sectors by using the BU and TD ensembles as documented in Extended Data Table 1 and Supplementary Table 2. Grouping GHG fluxes by sector may not precisely separate the contributions of human activities from natural components. For instance, wetland CH_4 emission is composed of a natural component (background emissions) and an anthropogenic contribution (for example, emissions altered by land use and climate change). Therefore, in this study, the anthropogenic contribution to the biogenic flux of each GHG is distinguished by removing modelled pre-industrial emissions from contemporary GHG estimates. To quantify the human-induced net biogenic balance of these three GHGs and its impact on the climate system, we use CO_2 equivalent units (CO_2 equiv.) based on the global warming potentials on a 100-year time horizon⁷ (GWP100; GWP defines the cumulative impacts that the emission of 1 g CH_4 or N_2O could have on the planetary energy budget relative to 1 g reference CO_2 gas over a certain period of years). This choice has been driven by the policy options being considered when dealing with biogenic GHG emissions and sinks^{7,11}. To address the changing relative importance of each gas as a function of the selected time frame, a supplementary calculation based on GWP metrics for a 20-year time horizon is also provided (GWP20; Table 1 and Methods).

We first examine the overall biogenic fluxes of all three gases in the terrestrial biosphere during the period 2001–10 ('the 2000s'; Fig. 1). The overall land biogenic CH_4 emissions estimated by TD and BU are very similar, $325 \pm 39 \text{ Tg C yr}^{-1}$ and $326 \pm 43 \text{ Tg C yr}^{-1}$ ($1 \text{ Tg} = 10^{12} \text{ g}$), respectively. Among the multiple land biogenic CH_4 sources (Extended Data Table 1), natural wetlands were the largest contributor, accounting for 40%–50% of total CH_4 emissions during the 2000s, while rice cultivation contributed about 10%. The remaining CH_4 emissions were from ruminants (~20%), landfills and waste (~14%), biomass burning (~4%–5%), manure management (~2%), and termites, wild animals and others (~6%–10%). Both TD and BU results suggest a global soil CH_4 sink that offsets approximately 10% of global biogenic CH_4 emissions, but this flux is poorly constrained, especially by atmospheric inversions, given its distributed nature and small magnitude.

¹International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, Auburn, Alabama 36849, USA. ²Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Iowa 50011, USA. ³Laboratoire des Sciences du Climat et de l'Environnement, 91191 Gif sur Yvette, France. ⁴Department of Global Ecology, Carnegie Institution for Science, Stanford, California 94305, USA. ⁵Global Carbon Project, CSIRO Oceans and Atmosphere Research, GPO Box 3023, Canberra, Australian Capital Territory 2601, Australia.

⁶Department of Environmental Sciences, Emory University, Atlanta, Georgia 30322, USA. ⁷School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, Arizona 86011, USA. ⁸School of Life Sciences, Arizona State University, Tempe, Arizona 85287, USA. ⁹College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK. ¹⁰NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, Colorado 80305, USA. ¹¹Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

¹²College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK. ¹³The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA. ¹⁴Institute of Ecosystems and Department of Ecology, Montana State University, Bozeman, Montana 59717, USA. ¹⁵Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ¹⁶Woods Hole Research Center, Falmouth, Massachusetts 02540, USA. ¹⁷Department of Earth and Planetary Science, Harvard University, 29 Oxford Street, Cambridge, Massachusetts 02138, USA.

Table 1 | Human-induced biogenic GHG emissions from the terrestrial biosphere based on GWP100 and GWP20 metrics

Metric	Human-induced GHG (\pm s.d.) (Pg CO ₂ equiv. yr ⁻¹)					
	1980s		1990s		2000s	
	TD	BU	TD	BU	TD	BU
GWP100						
CH ₄ source	7.5 (\pm 1.8)	7.9 (\pm 1.5)	7.4 (\pm 1.8)	6.9 (\pm 1.6)	7.4 (\pm 1.5)	7.5 (\pm 1.7)
N ₂ O source		2.8 (\pm 1.9)	1.6 (\pm 0.6)	2.9 (\pm 0.7)	2.2 (\pm 0.6)	3.3 (\pm 0.7)
CO ₂ sink	-1.4 (\pm 3.9)	-1.2 (\pm 4.0)	-3.2 (\pm 3.8)	-2.1 (\pm 4.1)	-5.8 (\pm 3.4)	-5.3 (\pm 4.5)
Overall GHG balance		9.4 (\pm 4.7)	5.9 (\pm 4.3)	7.7 (\pm 4.4)	3.9 (\pm 3.8)	5.4 (\pm 4.8)
Proportion of land CO ₂ sink being offset	-860%	-290%	-460%	-170%	-200%	
GWP20						
CH ₄ source	22.6 (\pm 5.4)	23.6 (\pm 4.6)	22.2 (\pm 5.5)	20.8 (\pm 4.7)	22.3 (\pm 4.6)	22.5 (\pm 5.1)
N ₂ O source		2.8 (\pm 1.9)	1.6 (\pm 0.6)	2.9 (\pm 0.7)	2.2 (\pm 0.6)	3.3 (\pm 0.7)
CO ₂ sink	-1.4 (\pm 3.9)	-1.2 (\pm 4.0)	-3.2 (\pm 3.8)	-2.1 (\pm 4.1)	-5.8 (\pm 3.4)	-5.3 (\pm 4.5)
Overall GHG balance		25.2 (\pm 6.4)	20.7 (\pm 6.7)	21.5 (\pm 6.3)	18.7 (\pm 5.8)	20.4 (\pm 6.8)
Proportion of land CO ₂ sink being offset	-2,120%	-760%	-1,110%	-430%	-480%	

Shown are estimated human-induced biogenic fluxes of CO₂, CH₄ and N₂O in the terrestrial biosphere for the 1980s, 1990s and 2000s based on global warming potential on 100-year and 20-year time horizons (GWP100 and GWP20, respectively). Numbers in parenthesis represent 1 s.d. (standard deviation). TD and BU stand for top-down and bottom-up estimates, respectively. The percentage numbers represent the proportion of land CO₂ sink that has been offset by human-induced CH₄ and N₂O emissions in the terrestrial biosphere. Detailed data sources and literature cited are provided in Supplementary Information.

Global biogenic N₂O emissions were estimated to be 12.6 ± 0.7 Tg N yr⁻¹ and 15.2 ± 1.0 Tg N yr⁻¹ by TD and BU methods, respectively. Natural ecosystems were a major source, contributing $\sim 55\%-60\%$ of all land biogenic N₂O emissions during the 2000s, the rest being from agricultural soils ($\sim 25\%-30\%$), biomass burning ($\sim 5\%$), indirect emissions ($\sim 5\%$), manure management ($\sim 2\%$), and human sewage ($\sim 2\%$).

The estimates of the global terrestrial CO₂ sink in the 2000s are -1.6 ± 0.9 petagrams of carbon a year (where $1 \text{ Pg} = 10^{15} \text{ g}$) (TD) and $-1.5 \pm 1.2 \text{ Pg C yr}^{-1}$ (BU). This estimate is comparable with the most recent estimates⁴, but incorporates more data sources (Supplementary Table 1).

Some CH₄ and N₂O emissions were present during pre-industrial times, while the global pre-industrial land CO₂ uptake was approximately in balance with the transport of carbon by rivers to the ocean and a compensatory ocean CO₂ source¹². Thus, the net land-atmosphere CO₂ flux reported here represents fluxes caused by human activities. In contrast, for CH₄ and N₂O only the difference between current and pre-industrial emissions represents net drivers of anthropogenic climate change. When subtracting modelled pre-industrial biogenic CH₄ and N₂O emissions of $125 \pm 14 \text{ Tg C yr}^{-1}$ and $7.4 \pm 1.3 \text{ Tg N yr}^{-1}$, respectively, from the contemporary estimates (see Methods), we find

the heating capacity of human-induced land biogenic CH₄ and N₂O emissions is opposite in sign and equivalent in magnitude to 1.7 (TD) and 2.0 (BU) times that of the current (2000s) global land CO₂ sink using 100-year GWPs (Fig. 1, Table 1). Hence there is a net positive cumulative impact of the three GHGs on the planetary energy budget, with our ‘best estimate’ being $3.9 \pm 3.8 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$ (TD) and $5.4 \pm 4.8 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$ (BU).

An alternative GWP metric (for example, GWP20 instead of GWP100) changes the relative importance of each gas, and gives a different view of the potential of various mitigation options¹¹. Using GWP20 values, the radiative forcing of contemporary (2000s) human-induced biogenic CH₄ emission alone is 3.8 (TD) or 4.2 (BU) times that of the land CO₂ sink in magnitude but opposite in sign, much larger than its role using the GWP100 metric (Table 1). Therefore, cutting CH₄ emissions is an effective pathway for rapidly reducing GHG-induced radiative forcing and the rate of climate warming in a short time frame^{8,11}.

On a 100-year time horizon, the cumulative radiative forcing of agricultural and waste emissions alone, including CH₄ from paddy fields, manure management, ruminants, and landfill and waste, along with N₂O emissions from crop cultivation, manure management, human

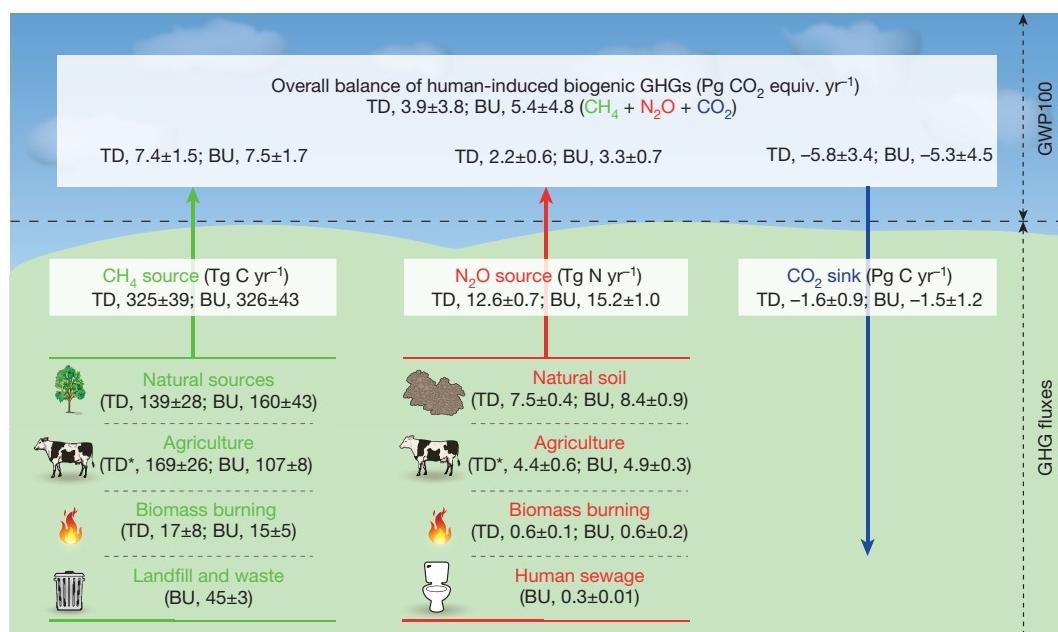


Figure 1 | The overall biogenic GHG balance of the terrestrial biosphere in the 2000s. Top-down (TD) and bottom-up (BU) approaches are used to estimate land CO₂ sink, CH₄ and N₂O fluxes for four major categories merged from 14 sectors (Extended Data Table 1). Global warming potential (GWP100) is calculated after removing pre-industrial biogenic emissions of CH₄ ($125 \pm 14 \text{ Tg C yr}^{-1}$) and N₂O ($7.4 \pm 1.3 \text{ Tg N yr}^{-1}$). Negative values indicate GHG sinks and positive values indicate GHG sources. TD* indicates estimates of agricultural CH₄ and N₂O emissions that include CH₄ sources from landfill and waste, and an N₂O source from human sewage, respectively.

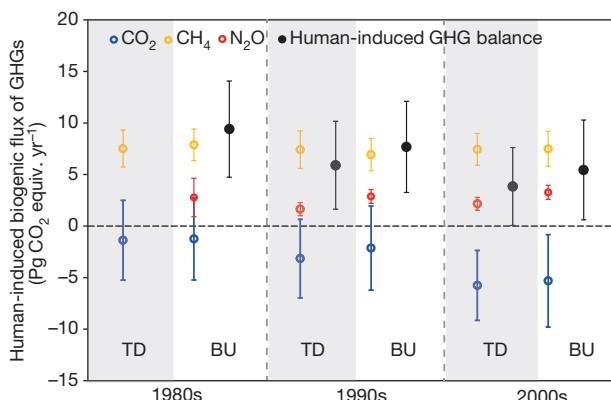


Figure 2 | Changes in the decadal balance of human-induced biogenic GHGs in the past three decades (based on GWP100). Data points show individual gases (blue for CO₂, yellow for CH₄, and red for N₂O) and net human-induced GHG balance (black) derived from biogenic sources with pre-industrial biogenic CO₂ sink, and CH₄ and N₂O emissions removed. Error bars, \pm s.d. calculated from various estimate ensembles.

sewage and indirect emissions, are estimated to be 7.9 ± 0.5 Pg CO₂ equiv. yr⁻¹ (BU) and 8.2 ± 1.0 Pg CO₂ equiv. yr⁻¹ (TD) for the 2000s, offsetting the human-induced land CO₂ sink by 1.4 to 1.5 times, respectively. In other words, agriculture and waste are the largest contributor to this twofold offset of the land CO₂ sink.

We further examine the change of human-induced biogenic GHG fluxes over past three decades (Fig. 2, Table 1). The net biogenic GHG source shows a decreasing trend of 2.0 Pg CO₂ equiv. yr⁻¹ per decade ($P < 0.05$), primarily due to an increased CO₂ sink—(2.2 Pg CO₂ equiv. yr⁻¹ per decade (TD) and 2.0 Pg CO₂ equiv. yr⁻¹ per decade (BU), $P < 0.05$)—as driven by a combination of increasing atmospheric CO₂ concentrations, forest regrowth, and nitrogen deposition³. The net emissions of CO₂ from tropical deforestation, included in the above net land CO₂ sink estimates, were found to decline or remain stable owing to reduced deforestation and increased forest regrowth¹³. However, one recent study based on

satellite observations¹⁴ suggests that the decreased deforestation in Brazil has been offset by an increase in deforestation in other tropical countries during 2000–12.

There is no clear decadal trend in total global biogenic CH₄ emissions from 1981 to 2010⁵. Since 2007, increased CH₄ emissions seem to result in a renewed and sustained increase of atmospheric CH₄, although the relative contribution of anthropogenic and natural sources is still uncertain^{15–17}. The BU estimates suggest an increase in human-induced biogenic N₂O emissions since 1980, at a rate of 0.25 Pg CO₂ equiv. yr⁻¹ per decade ($P < 0.05$), mainly due to increasing nitrogen deposition and nitrogen fertilizer use, as well as climate warming¹⁸. With pre-industrial emissions removed, the available TD estimates of N₂O emissions during 1995–2008 reflect a similar positive trend, although they cover a shorter period¹⁹.

The human-induced biogenic GHG fluxes vary by region (Fig. 3). Both TD and BU approaches indicate that human-caused biogenic fluxes of CO₂, CH₄ and N₂O in the biosphere of Southern Asia (Fig. 3) led to a large net climate warming effect, because the 100-year cumulative effects of CH₄ and N₂O emissions together exceed that of the terrestrial CO₂ sink. Southern Asia has about 90% of the global rice fields²⁰ and represents more than 60% of the world's nitrogen fertilizer consumption²¹, with 64%–81% of CH₄ emissions and 36%–52% of N₂O emissions derived from the agriculture and waste sectors (Supplementary Table 3). Given the large footprint of agriculture in Southern Asia, improved fertilizer use efficiency, rice management and animal diets could substantially reduce global agricultural N₂O and CH₄ emissions^{22,23}.

Africa is estimated to be a small terrestrial biogenic CO₂ sink (BU) or a CO₂-neutral region (TD), but it slightly warms the planet when accounting for human-induced biogenic emissions of CH₄ and N₂O, which is consistent with the finding of a recent study²⁴. South America is estimated to be neutral or a small sink of human-induced biogenic GHGs, because most current CH₄ and N₂O emissions in this region were already present during the pre-industrial period, and therefore do not represent new emissions since the pre-industrial era. Using the GWP100 metric, CO₂ uptake in North America and Northern Asia is almost equivalent in magnitude or even larger

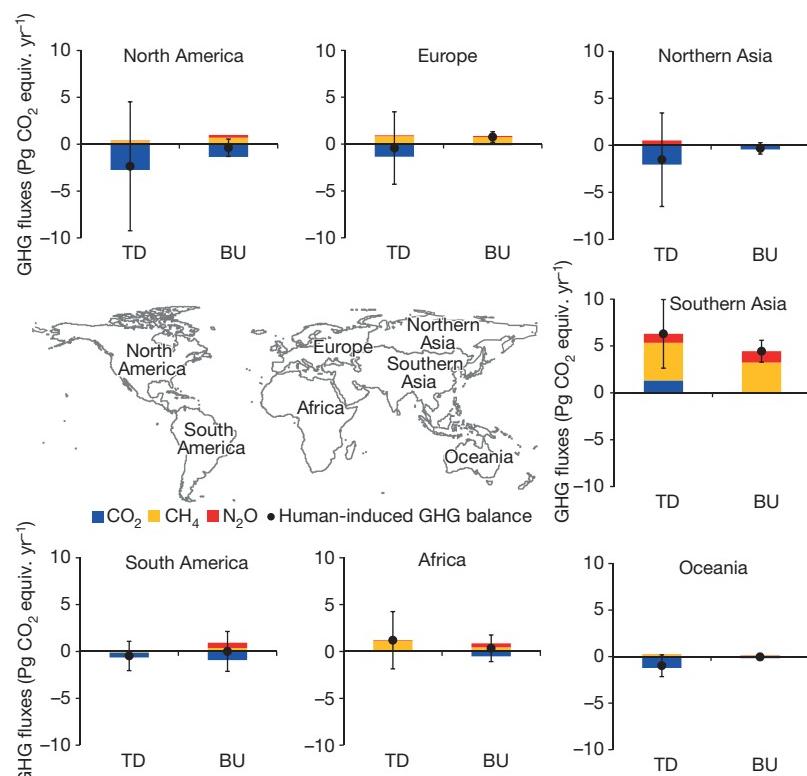


Figure 3 | The balance of human-induced biogenic GHGs for different continents in the 2000s (based on GWP100). Blue bars represent CO₂ flux, yellow CH₄ flux and red for N₂O flux with pre-industrial fluxes removed. Black dots indicate net human-induced GHG balance; error bars, \pm s.d. of estimate ensembles.

than human-caused biogenic CH₄ and N₂O emissions but opposite in sign, implying a small but significant role of the land biosphere in mitigating climate warming. Europe's land ecosystem is found to play a neutral role, similar to a previous synthesis study⁹ using both BU and TD approaches.

Compared to global estimations, much more work on regional GHG budgets is needed^{18,19}, particularly for tropical areas, as large uncertainty is revealed in both TD- and BU-derived GHG estimations. TD methods are subject to large uncertainties in their regional attribution of GHG fluxes to different types of sources. Furthermore, some TD estimates used BU values as priors, and may be heavily influenced by these assumed priors in regions where atmospheric observations are sparse. In contrast, BU approaches are able to consider region-specific disturbances and drivers (for example, insects and disease outbreaks) that are important at regional scale but negligible at global scale. However, the shortcoming of BU estimates is that they may not be consistent with the well-observed global atmospheric growth rates of GHGs. Also, accurate BU assessments are hindered by our limited understanding of microbial and below-ground processes and the lack of spatially-explicit, time-series data sets of drivers (for example, wildfire, peatland drainage, wetland extent). The magnitudes of human-induced CH₄ and N₂O emissions reported here are more uncertain than the total emissions of these gases because they contain the uncertainty of both pre-industrial emission and contemporary emission estimates (see Methods for additional discussion).

This study highlights the importance of including all three major GHGs in global and regional climate impact assessments, mitigation option and climate policy development. We should be aware of the likely countervailing impacts of mitigation efforts, such as enhanced N₂O emissions with soil carbon sequestration²⁵, increased CO₂ and N₂O emissions with paddy-drying to reduce CH₄ emissions²⁶, enhanced CH₄ emissions with peatland fire suppression and rewetting to reduce CO₂ and N₂O emissions²⁷, and increased indirect emissions from biofuel production²⁸. The future role of the biosphere as a source or sink of GHGs will depend on future land-use intensification pathways and on the evolution of the land CO₂ sinks²⁹. If the latter continues to increase as observed in the past three decades⁴, the overall biospheric GHG balance could be reversed. However, the evolution of the land CO₂ sink remains uncertain, with some projections showing an increasing sink in the coming decades³, while others showing a weakening sink due to the saturation of the CO₂ fertilization effect and positive carbon-climate feedbacks^{3,30}. Increasing land-use intensification using today's practices to meet food and energy demands will probably increase anthropogenic GHG emissions²³. However, the results of this study suggest that adoption of best practices to reduce GHG emissions from human-impacted land ecosystems could reverse the biosphere's current warming role.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 January; accepted 9 December 2015.

1. Lovelock, J. E. & Margulis, L. Atmospheric homeostasis by and for the biosphere: the gaia hypothesis. *Tellus A* **26**, <http://dx.doi.org/10.3402/tellusa.v26i1.29731> (1974).
2. Vitousek, P. M., Mooney, H. A., Lubchenco, J. & Melillo, J. M. Human domination of Earth's ecosystems. *Science* **277**, 494–499 (1997).
3. Ciais, P. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 6 (Cambridge Univ. Press, 2013).
4. Le Quéré, C. et al. Global carbon budget 2013. *Earth Syst. Sci. Data* **6**, 235–263 (2014).
5. Kirschke, S. et al. Three decades of global methane sources and sinks. *Nature Geosci.* **6**, 813–823 (2013).
6. Davidson, E. A. & Kanter, D. Inventories and scenarios of nitrous oxide emissions. *Environ. Res. Lett.* **9**, 105012 (2014).
7. Myhre, G. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 8 (Cambridge Univ. Press, 2013).
8. Montzka, S., Dlugokencky, E. & Butler, J. Non-CO₂ greenhouse gases and climate change. *Nature* **476**, 43–50 (2011).

9. Schulze, E. et al. Importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance. *Nature Geosci.* **2**, 842–850 (2009).
10. Tian, H. et al. North American terrestrial CO₂ uptake largely offset by CH₄ and N₂O emissions: toward a full accounting of the greenhouse gas budget. *Clim. Change* **129**, 413–426 (2015).
11. Allen, M. *Short-lived Promise? The Science and Policy of Cumulative and Short-Lived Climate Pollutants* <http://www.oxfordmartin.ox.ac.uk/publications/view/1960> (Oxford Martin Policy Paper, Oxford Martin School, Univ. Oxford, 2015).
12. Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L. & Gloos, M. A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes. *Glob. Biogeochem. Cycles* **21**, GB1019 (2007).
13. Tubiello, F. N. et al. The contribution of agriculture, forestry and other land use activities to global warming, 1990–2012. *Glob. Change Biol.* **21**, 2655–2660 (2015).
14. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
15. Dlugokencky, E. et al. Observational constraints on recent increases in the atmospheric CH₄ burden. *Geophys. Res. Lett.* **36**, L18803 (2009).
16. Rigby, M. et al. Renewed growth of atmospheric methane. *Geophys. Res. Lett.* **35**, L22805, <http://dx.doi.org/10.1029/2008GL036037> (2008).
17. Nisbet, E. G., Dlugokencky, E. J. & Bousquet, P. Methane on the rise — again. *Science* **343**, 493–495 (2014).
18. Tian, H. et al. Global methane and nitrous oxide emissions from terrestrial ecosystems due to multiple environmental changes. *Ecosyst. Health Sustain.* **1**, 4 (2015).
19. Saikawa, E. et al. Global and regional emissions estimates for N₂O. *Atmos. Chem. Phys.* **14**, 4617–4641 (2014).
20. Yan, X., Akiyama, H., Yagi, K. & Akimoto, H. Global estimations of the inventory and mitigation potential of methane emissions from rice cultivation conducted using the 2006 Intergovernmental Panel on Climate Change Guidelines. *Glob. Biogeochem. Cycles* **23**, GB2002 (2009).
21. FAO. *Current World Fertilizer Trends and Outlook to 2015* <http://www.fao.org/3/aav252e.pdf> (Food and Agriculture Organization of the United Nations, Rome, 2011).
22. Smith, P. et al. in *Climate Change 2014: Mitigation of Climate Change* (eds Edenhofer, O. et al.) Ch. 11, 816–922 (Cambridge Univ. Press, 2014).
23. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA* **108**, 20260–20264 (2011).
24. Valentini, R. et al. A full greenhouse gases budget of Africa: synthesis, uncertainties, and vulnerabilities. *Biogeosciences* **11**, 381–407 (2014).
25. Li, C., Frolking, S. & Butterbach-Bahl, K. Carbon sequestration in arable soils is likely to increase nitrous oxide emissions, offsetting reductions in climate radiative forcing. *Clim. Change* **72**, 321–338 (2005).
26. Yu, K., Chen, G. & Patrick, W. H. Reduction of global warming potential contribution from a rice field by irrigation, organic matter, and fertilizer management. *Glob. Biogeochem. Cycles* **18**, GB3018 (2004).
27. Murdiyarso, D., Hergoualc'h, K. & Verchot, L. Opportunities for reducing greenhouse gas emissions in tropical peatlands. *Proc. Natl. Acad. Sci. USA* **107**, 19655–19660 (2010).
28. Melillo, J. M. et al. Indirect emissions from biofuels: how important? *Science* **326**, 1397–1399 (2009).
29. Canadell, J. G. & Schulze, E. D. Global potential of biospheric carbon management for climate mitigation. *Nature Commun.* **5**, 5282 (2014).
30. Stocker, B. D. et al. Multiple greenhouse-gas feedbacks from the land biosphere under future climate change scenarios. *Nature Clim. Change* **3**, 666–672 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported partially by NASA grants (NNX08AL73G, NNX14AO73G, NNX10AU06G, NNX11AD47G, NNG04GM39C) and NSF grants (AGS 1243232, AGS-1243220, CNH1210360). J.G.C. was supported by the Australian Climate Change Science Program. E.S. was supported by the NOAA Climate Program Office (award NA13OAR4310059). C.R.S. was supported by NASA grants (NNX12AP74G, NNX10AQ01A, NNX11AO08A). K.R.G. was supported by NSF CAREER (AGS-0846358). R.G.P. was supported by a NASA Upper Atmosphere Research Program AGAGE grant (NNX11AF17G to MIT). This study contributes to the Non-CO₂ Greenhouse Gases Synthesis of NACP (North American Carbon Program), and the Global Carbon Project (a joint project of IGBP, IHDP, WCRP and Diversitas).

Author Contributions H.T. initiated this research and was responsible for the integrity of the work as a whole. H.T. and C.L. performed analysis, calculations and drafted the manuscript. P.C., A.M.M. and J.G.C. contributed to data synthesis and manuscript development. B.Z., J.Y., G.C. and S.P. contributed to data collection and analysis. E.S., D.N.H., K.R.G., S.S., P.B., L.B., E.D., P.F., J.M., B.P., R.G.P., M.S., C.R.S. and S.C.W. contributed to data provision, data processing, or interpretation. All authors discussed and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.T. (tianhan@auburn.edu or tian.hanqin@gmail.com).

METHODS

Definition of biogenic GHG fluxes. In this study, we define land biogenic GHG fluxes as those originating from plants, animals, and microbial communities, with changes driven by both natural and anthropogenic perturbations. For example, this analysis considers the biosphere–atmosphere CO₂ flux resulting from the direct and indirect effects of anthropogenic activities, such as land use and management, climate warming, rising atmospheric CO₂, and nitrogen deposition, but excludes CO₂ emissions due to geological processes (for example, volcanic eruption, weathering), fossil fuel combustion, and cement production. Biogenic CH₄ fluxes include land–atmosphere CH₄ emissions by natural wetlands, rice cultivation, biomass burning, manure management, ruminants, termites, landfills and waste, as well as soil CH₄ uptake. Biogenic N₂O emissions include those released from agricultural ecosystems (that is, fertilized soil emission, manure management, and indirect N₂O emission from manure and synthetic nitrogen fertilizer use), natural ecosystems (that is, soil emissions and emissions from nitrogen re-deposition), human sewage, and biomass burning.

Data sources and calculation. We synthesized estimates of biogenic CO₂, CH₄ and N₂O fluxes in the terrestrial biosphere derived from 28 bottom-up (BU) studies and 13 top-down (TD) atmospheric inversion studies for two spatiotemporal domains (global scale during 1981–2010 and continental scale during the 2000s). First, the data we compiled include the most recent estimates of individual GHG gases from multi-model inter-comparison projects (for example, Atmospheric Tracer Transport Model Inter-comparison Project—TransCom³¹, Trends in net land atmosphere carbon exchanges—TRENDY³², and Multi-scale Synthesis and Terrestrial Model Inter-comparison Project—MsTMIP³³). Second, the estimate ensembles included the published global synthesis results that report decadal land-atmosphere GHG exchange during 1981–2010^{4–6}. Third, for those items that lack detailed information from the above estimations (for example, continental estimate of CH₄ emission from rice fields and soil CH₄ sink, Supplementary Table 1), we use multi-source published estimates and a recent process-based modelling result¹⁸. We limit literature reporting the continental GHG estimate to those studies that have close boundary delineation with our definition, and that have gas flux estimates covering all continents. Only part of the global studies we used has provided continental estimates (details on data sources can be found in Supplementary Table 1 and Supplementary Information Section S3).

In Le Quéré *et al.*⁴, net land CO₂ flux is the sum of carbon emission due to land-use change (E_{LUC}) and the residual terrestrial carbon sink (S_{LAND}). Estimates of budget residual, as one of the top-down approaches, are calculated as the sum of E_{LUC} and S_{LAND} (cited from table 7 of Le Quéré *et al.*⁴). Land CO₂ sink estimated by the TRENDY model inter-comparison project³² does not account for land-use effects on terrestrial carbon dynamics, and we therefore add land-use-induced carbon fluxes as estimated by IPCC AR5³ (table 6.3) to obtain the net land carbon sink estimates. However, the land CO₂ sink estimated by the MsTMIP project³⁴ is derived from model simulations considering climate variability, atmospheric CO₂ concentration, nitrogen deposition, as well as land-use change. We directly use its model ensemble estimates in this study. In addition, BU estimates of land CO₂ sink^{4,34} have been adjusted by removing the CO₂ emissions from drained peatland globally^{13,35}, because global land ecosystem models usually overlook this part of carbon loss.

We include TD and BU estimates of CH₄ and N₂O emission from biomass burning. The TD approach (for example, CarbonTracker-CH₄, Bruhwiler *et al.*³⁶) considers all the emission sources and growth rate in atmospheric concentration. For BU estimation (for example, DLEM simulation, Tian *et al.*³⁷), they use historical fire data that is developed from satellite images and the historical record, to drive a process-based land ecosystem model, so the change in fire occurrence is naturally considered. Other BU estimates, for example, GFED (Van der Werf *et al.*³⁸) and EDGAR³⁹, all include peatland fire emissions. We remove pre-industrial CH₄ and N₂O emission that includes sources from biomass burning to estimate human-caused gas fluxes in the terrestrial biosphere. The role of peatland fire in the estimated CO₂ flux is similar to CH₄ and N₂O estimation: fire emission is included in the TD approach and historical fire is included as one of input drivers (or counted as part of land-use change in most BU models—for example, fire occurrence in deforestation and cropland expansion) in some models. Although peatland fire emission caused by human activities is counted in our analysis, like other sectors, we cannot distinguish how much peat fire is caused by human activity since no specific information is available on pre-industrial peatland fire emission.

In summary, this study provides multi-level estimates on biogenic GHG fluxes, including global biogenic fluxes of CO₂, CH₄, and N₂O during 1981–2010, continental-level estimates on biogenic fluxes of CO₂, CH₄ and N₂O over the 2000s, and sector-based estimates on biogenic CH₄ and N₂O fluxes over the 2000s. Extended Data Table 1 shows our estimates of biogenic CH₄ fluxes for 8 sectors and N₂O fluxes for 6 sectors. These sectors are further merged into four major categories for CH₄ and N₂O fluxes, respectively (Fig. 1).

All the raw data and relevant calculations can be found in Supplementary Table 2. Human-induced biogenic CH₄ and N₂O emissions are calculated by subtracting the pre-industrial emissions as estimated below.

Pre-industrial biogenic GHG estimations. Here we provide a description of how we estimated the pre-industrial GHG emissions. For CO₂ flux, since terrestrial ecosystem models assume the net land–air carbon flux in the pre-industrial era is zero and the modelled C sink is solely human-driven, in order to make TD estimates comparable to BU estimates, the CO₂ sink from TransCom simulations³¹ has been adjusted by removing the natural CO₂ sink (0.45 Pg C yr⁻¹)¹² due to riverine transport from land to ocean. This CO₂ sink of 0.45 Pg C yr⁻¹ was allocated to each continent by using continental-scale estimates of riverine carbon export by Ludwig *et al.*⁴⁰ and assuming 100 Tg C yr⁻¹ of organic carbon is buried and 50% of DIC (dissolved inorganic carbon) export is degassing⁴¹.

Human-induced biogenic CH₄ and N₂O emissions are calculated by subtracting the pre-industrial emissions. We define pre-industrial emissions as the GHG source under pre-industrial environmental conditions and land-use patterns, including CH₄ and N₂O emissions from land ecosystems (for example, natural wetlands, forests, grassland, shrublands). The pre-industrial CH₄ estimate (125.4 ± 14.4 Tg C yr⁻¹) is composed of CH₄ emission from natural wetland and vegetation (99.2 ± 14.3 Tg C yr⁻¹ derived from Houweling *et al.*⁴², Basu *et al.*⁴³ and an unpublished result (H.T.) from DLEM model simulation with potential vegetation map (excluding cropland cultivation and other anthropogenic activities)), termites (15 Tg C yr⁻¹, Slugokenky *et al.*⁴⁴), and wildfire and wild animals (3.75–7.5 Tg C yr⁻¹ each, Slugokenky *et al.*⁴⁴). Pre-industrial N₂O emission (7.4 ± 1.3 Tg N yr⁻¹) is derived from the estimate of terrestrial N₂O emission (6.6 ± 1.4 Tg N yr⁻¹) by Davidson and Kanter⁶, and a DLEM simulation (H.T., unpublished results) (8.1 ± 1.2 Tg N yr⁻¹) driven by environmental factors at pre-industrial level and a potential vegetation map.

Calculation and interpretation of GWP. GWP is used to define the cumulative impacts that the emission of 1 g CH₄ or N₂O could have on the planetary energy budget relative to 1 g reference CO₂ gas over a certain period (for example, GWP100 and GWP20 for 100 or 20 years). To calculate CO₂ equivalents of the human-induced biogenic GHG balance, we adopt 100-year GWPs of 28 and 265 for CH₄ and N₂O, respectively, and 20-year GWPs of 84 and 264, respectively⁷. These values of GWP20 and GWP100 used in this study do not include carbon–climate feedbacks. The different contributions of each gas to the net GHG balance will vary using different GWP time horizons (for example, GWP20 versus GWP100, see Table 1). In this study, we applied the following equation to calculate the human-induced biogenic GHG balance:

$$\text{GHG} = F_{\text{CO}_2-\text{C}} \frac{44}{12} + F_{\text{CH}_4-\text{C}} \frac{16}{12} \times \text{GWP}_{\text{CH}_4} + F_{\text{N}_2\text{O}-\text{N}} \frac{44}{28} \times \text{GWP}_{\text{N}_2\text{O}}$$

Where $F_{\text{CO}_2-\text{C}}$, $F_{\text{CH}_4-\text{C}}$ and $F_{\text{N}_2\text{O}-\text{N}}$ are annual exchanges (unit: Pg C yr⁻¹ or Pg N yr⁻¹) of human-induced biogenic CO₂, CH₄ and N₂O between terrestrial ecosystems and the atmosphere based on the mass of C and N, respectively. The fractions 44/12, 16/12 and 44/28 were used to convert the mass of CO₂-C, CH₄-C and N₂O-N into CO₂, CH₄ and N₂O. GWP_{CH₄} (Pg CO₂ equiv. per Pg CH₄) and GWP_{N₂O} (Pg CO₂ equiv. per Pg N₂O) are constants indicating integrated radiative forcing of CH₄ and N₂O in terms of a CO₂ equivalent unit.

Nevertheless, it is noted that adoption of GWP100 to calculate CO₂ equivalent is not fundamentally scientific but depends on a policy perspective. The relative importance of each gas at a certain time period and likely mitigation option could change due to GWP metrics at different time horizon (for example, GWP20 and GWP100 according to Myhre *et al.*⁷, Table 1). For example, CH₄ has a shorter lifetime (~9 years), and its cumulative radiative forcing is equivalent to 84 times the same amount of CO₂ over 20 years, and 28 times the same amount of CO₂ over 100 years. At a 20-year time horizon, anthropogenic CH₄ and N₂O emissions in the 2000s are equivalent to 4.2–4.8 (TD-BU) times the land CO₂ sink in magnitude but opposite in sign, and the net balance of human-induced GHG in the terrestrial biosphere is 20.4 ± 6.8 Pg CO₂ equiv. yr⁻¹ and 18.7 ± 5.8 Pg CO₂ equiv. yr⁻¹ as estimated by BU and TD approaches, respectively. Among them, anthropogenic CH₄ emissions are 7–10 times (BU-TD) as much as N₂O emissions in terms of GWP20. At a 20-year time horizon, the cumulative radiative forcing of contemporary anthropogenic CH₄ emission alone is 3.8–4.2 (TD-BU) times as much as that of the land CO₂ sink but opposite in sign, larger than its role at a 100-year time horizon (1.3–1.4 times the radiative forcing of the CO₂ sink). Therefore, to cut CH₄ emission could rapidly reduce GHG-induced radiative forcing in a short time frame^{7,8,44}.

Statistics. We use mean ± 1 standard deviation (s.d.) to indicate the best estimates and their ranges. Estimate ensembles are grouped for the TD and BU approaches, and the mean value of multiple ensembles is calculated for each gas in a certain region and period. In the TD and BU groups, we assume the individual estimates are independent of each other, and therefore, the s.d. for each ensemble

mean is calculated as the square root of the quadratic sum of s.d.s reported in each estimate.

Sample size. No statistical methods were used to predetermine sample size.

Continental-level estimations and divergence of biogenic-GHG fluxes. Using the TD and BU ensembles, we estimated the net human-induced biogenic GHG balance during the 2000s for 7 continents or regions, which include North America, South America, Europe, Northern Asia, Southern Asia, Africa and Oceania (Fig. 3). Primarily owing to large CH₄ and N₂O emissions, both approaches show that Southern Asia is a net human-induced biogenic GHG source of $6.3 \pm 3.7 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$ and $4.4 \pm 1.2 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$ as estimated by TD and BU, respectively, with the GWP100 metric (Supplementary Table 3). Southern Asia has about 90% of the global rice fields and represents over 60% of the world's nitrogen fertilizer consumption. China and India together consume half of the global nitrogen fertilizer²¹. This leads to the highest regional CH₄ and N₂O emissions, as the two approaches consistently reveal. This finding is also consistent with previous studies conducted in China and India^{45–47}. South America was estimated to be a CO₂ sink with a large uncertainty (Supplementary Table 3). Although South America is a large CH₄ and N₂O source, most of these emissions are present at pre-industrial times. Natural wetlands in South America accounted for 31%–40% of global wetland CH₄ emissions in the 2000s, and 26%–30% of the global natural soil N₂O emissions were derived from this region. Therefore, the contribution of this continent to human-induced GHG balance is negligible or acts as a small sink. Likewise, Africa is estimated to be a small CO₂ sink or a CO₂-neutral region, but adding CH₄ and N₂O emissions makes this continent contribute a small positive radiative forcing, slightly warming the planet.

North America and Northern Asia are found to be a neutral region to a net human-induced biogenic GHG sink, with 100-year cumulative radiative forcing of biogenic CH₄ and N₂O emissions fully or partially offsetting that of the land CO₂ sink in this continent (Supplementary Table 3). The largest CO₂ sink was found in North America, ranging from $-0.37 \pm 0.22 \text{ Pg C yr}^{-1}$ to $-0.75 \pm 1.87 \text{ Pg C yr}^{-1}$ as estimated by BU and TD, respectively, probably due to a larger area of highly productive and intensively managed ecosystems (for example, forests, woodlands, and pasture) that were capable of sequestering more CO₂. Our estimate falls within the newly-reported CO₂ sink of -0.28 to $-0.89 \text{ Pg C yr}^{-1}$ in North America obtained by synthesizing inventory, atmospheric inversions, and terrestrial modelling estimates⁴⁸. Considering the three gases together, TD estimates showed that North America acts as a net GHG sink with a large s.d. (human-induced biogenic GHG of $-2.35 \pm 6.87 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$, Fig. 3 and Supplementary Table 3). By contrast, BU estimates suggested that North America was a small GHG sink of $-0.38 \pm 0.93 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$ based on GWP100. Our estimate is comparable to previous GHG budget syntheses for North America^{10,37}. TD estimates indicated that Oceania and Europe act to give a small negative net radiative forcing over 100 years (-0.98 ± 1.17 and $-0.42 \pm 3.86 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$, respectively), while BU estimates indicated a negligible contribution in Oceania, and a positive net radiative forcing ($0.76 \pm 0.57 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$) in Europe. According to BU estimates, CO₂ emission from drained peatland in Europe accounted for about one-third of the global total during the 2000s³⁵, which partially explains the warming effect of biogenic GHG in this region as revealed by BU.

It is important to note that only human-caused biogenic GHG fluxes are included in this study, and the regional GHG balance will clearly move towards a net source if the emissions related to fossil fuel combustion and usage are taken into account.

Our analyses indicate that the TD and BU estimates show a larger divergence at the continental scale than at the global scale. We note that the high radiative forcing estimate of human-induced biogenic GHG balance ($6.30 \pm 3.66 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$) in the TD approach in Southern Asia is partially because the land biosphere in this region is estimated to be a net CO₂ source of $0.36 \text{ Pg C yr}^{-1}$ with a large s.d. of $0.99 \text{ Pg C yr}^{-1}$ by TransCom inversions^{31,49}. It includes CO₂ sources and sinks from respiration, primary production, disturbances, rivers outgassing, and land-use change. In contrast, most BU estimations using land ecosystem models do not consider the full set of factors responsible for CO₂ release^{32,33}. The discrepancy between TD and BU estimates for Southern Asia may occur for several reasons. First, the land-use history data commonly used for driving terrestrial biosphere models, for example, HYDE⁵⁰ and GLM⁵¹, were reported to overestimate cropland area and cropland expansion rate in China and to under-estimate it in India compared to the regional data set^{52,53}, thus biasing BU estimates of land conversion-induced carbon fluxes. But none of the BU models included in this study conducted global simulation with such a regional data set updated. Second, large uncertainties exist in estimating carbon release due to tropical deforestation^{4,54–57}. Third, carbon emissions due to peat fires and peatland drainage were a large but usually ignored carbon source in tropical Asia (EDGAR 4.2³⁹ and Joosten *et al.*³⁵). In the BU estimates we included, some models consider peat fire by using an input driver of fire regime from satellite images, while most of them do not consider drained

peatland and accelerated SOC (soil organic carbon) decomposition. Therefore, BU models may underestimate the CO₂ emissions from intensively-disturbed areas, resulting in a small CO₂ source of $0.03 \pm 0.29 \text{ Pg C yr}^{-1}$. BU estimations show that the net human-induced biogenic GHG balance in Southern Asia turned out to warm the planet with a 100-year cumulative radiative forcing of $4.44 \pm 1.17 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$.

Net GHG balance in Africa was positive but with a discrepancy between the TD and BU approaches. TD estimates suggested that Africa was a weak source of CO₂ and a strong source of CH₄ and N₂O, resulting in a positive net radiative forcing of $1.20 \pm 3.05 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$. However, BU ensembles estimated that African terrestrial biosphere acted as a relatively smaller climate warmer ($0.34 \pm 1.42 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$) due to an anthropogenic land sink of CO₂ ($-0.52 \pm 1.38 \text{ Pg CO}_2 \text{ equiv. yr}^{-1}$) and a strong source of CH₄ and N₂O. These divergent estimates in Africa occur for several reasons. First, it was difficult to constrain emissions using TD in this region, due to the lack of atmospheric data. No tropical continent is covered by enough atmospheric GHG measurement stations, making the TD results uncertain in those regions, with almost no uncertainty reduction from the prior knowledge assumed before inversion. Second, there were also large uncertainties in BU estimates. Some of the BU models ignored fire disturbance that is likely to result in a carbon source of $1.03 \pm 0.22 \text{ Pg C yr}^{-1}$ in Africa^{24,38} and this emission has been partially offset by carbon uptake due to regrowth. Another reason might be the overestimated CO₂ fertilization effect, which could be limited by nutrient availability. Only a few BU models addressed interactive nutrient cycles in their simulation experiments³².

Uncertainty sources and future research needs. A wide variety of methods, such as statistical extrapolations, and process-based and inverse modelling, were applied to estimate CO₂, CH₄ and N₂O fluxes. TD methods are subject to large uncertainties in their regional attribution of GHG fluxes to different type of sources⁵⁸. BU approaches are however limited by our understanding of underlying mechanisms and the availability and quality of input data. In addition, the TD approach is dependent on BU estimates as prior knowledge, especially in the tropics where both uncertainties are very large.

For example, terrestrial CO₂ uptake estimates from process-based model ensembles in Africa, South America, and Southern Asia are larger than those from TD approaches, while smaller than TD estimates in North America, Europe, Oceania and Northern Asia (Fig. 3, Supplementary Table 3). The larger BU CO₂ sink estimate might be related to biased land-use history data, excluded fire emission and CO₂ release due to extreme disturbances such as insect outbreaks and windthrow^{24,32}. Another reason is the lack of fully-coupled carbon–nitrogen cycles in most BU models that overestimate the CO₂ fertilization effect particularly in regions of large biomass and large productivity^{59–61}. However, a larger CO₂ sink observed from tropical regrowth forests compared to intact forests⁵⁵ might be underestimated because few models are capable of capturing CO₂ uptake related to tropical secondary forest management and age structure. The post-disturbance and plantation-induced shift towards rapid carbon accumulation in young forests that were poorly or not represented in terrestrial ecosystem models might be one of the factors responsible for CO₂ sink underestimation as revealed by several studies conducted in mid- and high-latitudes^{62–64}. The modelled ecosystem responses to frequent occurrence of extreme climate events in BU studies are another uncertainty in estimating variations of the land CO₂ sink^{65,66}.

The estimates of terrestrial CH₄ fluxes remain largely uncertain. One major uncertainty in BU wetland CH₄ emission estimate is wetland areal extent data⁶⁷. Global inundated area extent was reported to decline by approximately 6% during 1993–2007 with the largest decrease in tropical and subtropical South America and South Asia⁶⁸. However, the majority of BU models failed either in capturing dynamic inundation area or in simulating inundation and saturated conditions. Tropical emissions, the dominant contributor for global wetland emission, are particularly difficult to quantify owing to sparse observations for both TD (atmospheric mixing ratios) and BU (flux measurements) approaches and large inter-annual, seasonal variability, and a long-term change in the inundation extent for the BU modelling approach^{5,36,68}. At high latitudes, current dynamic inundation data could not well represent permanent wetlands⁶⁷, most of which are occupied by peatland. Because of large soil carbon storage in peatlands, such areas are an important CH₄ source. In addition, a large divergence exists in the estimation of rice field CH₄ emissions (Supplementary Table 2). The estimated global CH₄ emissions from rice fields are sensitive to rice field area, management practices (for example, water regime, fertilizer application), and local climate and soil conditions that directly affect activities of methanotroph and methanogen^{20,69,70}. Models need better representation of CH₄ production and consumption processes modified by agricultural management, such as continuous flooding, irrigation with intermediate drainage, or rainfed⁷⁰.

Compared to CO₂ and CH₄, there have been fewer studies of global N₂O emissions. The TD approach is constrained by sparse or inconsistent measurements

of atmospheric N₂O mixing ratios^{19,71}. Decadal trends during 1981–2010 from BU approaches were primarily from two process-based models^{18,72}, instead of IPCC methodology based on the N₂O emission factors. The major uncertainty source, therefore, includes data characterizing spatiotemporal variation of reactive nitrogen enrichment, modelling schemes representing multiple nitrogen forms, transformation, and their interactions with other biogeochemical and hydrological cycles, as well as key parameters determining the sensitivity of N₂O emission to temperature, soil moisture, and availability of oxygen^{45,46,72–74}. A large divergence exists in the estimation of natural soil N₂O emission by inventory, empirical and process-based models, implying that our understanding of the processes and their controls remain uncertain^{18,72,75–77}. Tropical areas are the major contributors to large divergence. N₂O sources from tropical undisturbed wetland and drained wetland/peatland are likely to be underestimated⁷⁸.

31. Gurney, K. R., Baker, D., Rayner, P. & Denning, S. Interannual variations in continental-scale net carbon exchange and sensitivity to observing networks estimated from atmospheric CO₂ inversions for the period 1980 to 2005. *Glob. Biogeochem. Cycles* **22**, GB3025 (2008).
32. Sitch, S. *et al.* Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences* **12**, 653–679 (2015).
33. Huntzinger, D. *et al.* The North American carbon program multi-scale synthesis and terrestrial model intercomparison project—part 1: Overview and experimental design. *Geosci. Model Dev.* **6**, 2121–2133 (2013).
34. Schwalm, C. R. *et al.* Toward “optimal” integration of terrestrial biosphere models. *Geophys. Res. Lett.* **42**, 4418–4428 (2015).
35. Joosten, H. *The Global Peatland CO₂ Picture: Peatland Status and Drainage Related Emissions in All Countries of the World* http://www.wetlands.org/Portals/0/publications/Report/The%20Global%20Peatland%20CO2%20Picture_web%20Aug%202010.pdf (Greifswald University Wetlands International, Ede, 2010).
36. Bruhwiler, L. *et al.* CarbonTracker-CH₄: an assimilation system for estimating emissions of atmospheric methane. *Atmos. Chem. Phys.* **14**, 8269–8293 (2014).
37. Tian, H. *et al.* Contemporary and projected biogenic fluxes of methane and nitrous oxide in North American terrestrial ecosystems. *Front. Ecol. Environ.* **10**, 528–536 (2012).
38. van der Werf, G. R. *et al.* Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmos. Chem. Phys.* **10**, 11707–11735 (2010).
39. EDGAR Emission Database for Global Atmospheric Research (EDGAR) release version 4.2 <http://edgar.jrc.ec.europa.eu/overview.php?v=42> (2014); accessed 2 February 2016.
40. Ludwig, W., Amiotte-Suchet, P. & Probst, J. ISLSCP II atmospheric carbon dioxide consumption by continental erosion. <http://dx.doi.org/10.3334/ORNLDAC/1019> (ORNL DAAC, 2011).
41. Sarmiento, J. & Sundquist, E. Revised budget for the oceanic uptake of anthropogenic carbon dioxide. *Nature* **356**, 589–593 (1992).
42. Houweling, S., Van der Werf, G., Klein Goldewijk, K., Röckmann, T. & Aben, I. Early anthropogenic CH₄ emissions and the variation of CH₄ and ¹³CH₄ over the last millennium. *Glob. Biogeochem. Cycles* **22**, GB1002 (2008).
43. Basu, A. *et al.* Analysis of the global atmospheric methane budget using ECHAM-MOZ simulations for present-day, pre-industrial time and the Last Glacial Maximum. *Atmos. Chem. Phys. Discuss.* **14**, 3193–3230 (2014).
44. Dlugokencky, E. J., Nisbet, E. G., Fisher, R. & Lowry, D. Global atmospheric methane: budget, changes and dangers. *Phil. Trans. R. Soc. A* **369**, 2058–2072 (2011).
45. Lu, C. & Tian, H. Net greenhouse gas balance in response to nitrogen enrichment: perspectives from a coupled biogeochemical model. *Glob. Change Biol.* **19**, 571–588 (2013).
46. Tian, H. *et al.* Net exchanges of CO₂, CH₄, and N₂O between China's terrestrial ecosystems and the atmosphere and their contributions to global climate warming. *J. Geophys. Res.* **116**, G02011 (2011).
47. Banger, K. *et al.* Biosphere–atmosphere exchange of methane in India as influenced by multiple environmental changes during 1901–2010. *Atmos. Environ.* **119**, 192–200 (2015).
48. King, A. *et al.* North America's net terrestrial CO₂ exchange with the atmosphere 1990–2009. *Biogeosciences* **12**, 399–414 (2015).
49. Baker, D. *et al.* TransCom 3 inversion intercomparison: impact of transport model errors on the interannual variability of regional CO₂ fluxes, 1988–2003. *Glob. Biogeochem. Cycles* **20**, GB1002, <http://dx.doi.org/10.1029/2004GB002439> (2006).
50. Goldewijk, K. K. Estimating global land use change over the past 300 years: the HYDE database. *Glob. Biogeochem. Cycles* **15**, 417–433 (2001).
51. Hurtt, G. *et al.* The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands. *Glob. Change Biol.* **12**, 1208–1229 (2006).
52. Liu, M. & Tian, H. China's land cover and land use change from 1700 to 2005: estimations from high-resolution satellite data and historical archives. *Glob. Biogeochem. Cycles* **24**, GB3003 (2010).
53. Tian, H., Banger, K., Bo, T. & Dadhwal, V. K. History of land use in India during 1880–2010: large-scale land transformations reconstructed from satellite data and historical archives. *Global Planet. Change* **121**, 78–88 (2014).
54. Friedlingstein, P. *et al.* Update on CO₂ emissions. *Nature Geosci.* **3**, 811–812 (2010).
55. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
56. Harris, N. L. *et al.* Baseline map of carbon emissions from deforestation in tropical regions. *Science* **336**, 1573–1576 (2012).
57. Archer-Nicholls, S. *et al.* Characterising Brazilian biomass burning emissions using WRF-Chem with MOSAIC sectional aerosol. *Geosci. Model Dev.* **8**, 549–577 (2015).
58. Peylin, P. *et al.* Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions. *Biogeosciences* **10**, 6699–6720 (2013).
59. Thornton, P. E. *et al.* Carbon–nitrogen interactions regulate climate–carbon cycle feedbacks: results from an atmosphere–ocean general circulation model. *Biogeosciences* **6**, 2099–2120 (2009).
60. Sokolov, A. P. *et al.* Consequences of considering carbon–nitrogen interactions on the feedbacks between climate and the terrestrial carbon cycle. *J. Clim.* **21**, 3776–3796 (2008).
61. Zaehele, S., Friedlingstein, P. & Friend, A. D. Terrestrial nitrogen feedbacks may accelerate future climate change. *Geophys. Res. Lett.* **37**, L01401 (2010).
62. Goward, S. N. *et al.* Forest disturbance and North American carbon flux. *Eos Trans. AGU* **89**, 105–106 (2008).
63. Williams, C. A., Collatz, G. J., Masek, J. & Goward, S. N. Carbon consequences of forest disturbance and recovery across the conterminous United States. *Glob. Biogeochem. Cycles* **26**, GB1005 (2012).
64. Bellassen, V. *et al.* Reconstruction and attribution of the carbon sink of European forests between 1950 and 2000. *Glob. Change Biol.* **17**, 3274–3292 (2011).
65. Reichstein, M. *et al.* Climate extremes and the carbon cycle. *Nature* **500**, 287–295 (2013).
66. Zscheischler, J. *et al.* Impact of large-scale climate extremes on biospheric carbon fluxes: an intercomparison based on MsTMIP data. *Glob. Biogeochem. Cycles* **28**, 585–600 (2014).
67. Melton, J. *et al.* Present state of global wetland extent and wetland methane modelling: conclusions from a model intercomparison project (WETCHIMP). *Biogeosciences* **10**, 753–788 (2013).
68. Prigent, C. *et al.* Changes in land surface water dynamics since the 1990s and relation to population pressure. *Geophys. Res. Lett.* **39**, L08403 (2012).
69. Ren, W. *et al.* Spatial and temporal patterns of CO₂ and CH₄ fluxes in China's croplands in response to multifactor environmental changes. *Tellus B* **63**, 222–240 (2011).
70. Banger, K., Tian, H. & Lu, C. Do nitrogen fertilizers stimulate or inhibit methane emissions from rice fields? *Glob. Change Biol.* **18**, 3259–3267 (2012).
71. Reay, D. S. *et al.* Global agriculture and nitrous oxide emissions. *Nature Clim. Change* **2**, 410–416 (2012).
72. Saikawa, E., Schlosser, C. & Prinn, R. Global modeling of soil nitrous oxide emissions from natural processes. *Glob. Biogeochem. Cycles* **27**, 972–989 (2013).
73. Bouwman, A. *et al.* Global trends and uncertainties in terrestrial denitrification and N₂O emissions. *Phil. Trans. R. Soc. Lond. B* **368**, 20130112, <http://dx.doi.org/10.1098/rstb.2013.0112> (2013).
74. Butterbach-Bahl, K., Diaz-Pines, E. & Dannenmann, M. in *Handbook of Global Environmental Pollution Vol. 1, Global Environmental Change* 325–334 (Springer, 2014).
75. Potter, C. S., Matson, P. A., Vitousek, P. M. & Davidson, E. A. Process modeling of controls on nitrogen trace gas emissions from soils worldwide. *J. Geophys. Res. Atmos.* **101**, 1361–1377 (1996).
76. Xu, X., Tian, H. & Hui, D. Convergence in the relationship of CO₂ and N₂O exchanges between soil and atmosphere within terrestrial ecosystems. *Glob. Change Biol.* **14**, 1651–1660 (2008).
77. Zhuang, Q. *et al.* Response of global soil consumption of atmospheric methane to changes in atmospheric climate and nitrogen deposition. *Glob. Biogeochem. Cycles* **27**, 650–663 (2013).
78. Lienggaard, L. *et al.* Extreme emission of N₂O from tropical wetland soil (Pantanal, South America). *Front. Microbiol.* **3**, 1–13 (2012).

Extended Data Table 1 | Decadal estimates of global terrestrial biogenic CO₂, CH₄ and N₂O fluxes

GHG	Sector	1980s		1990s		2000s	
		Top-down	Bottom-up	Top-down	Bottom-up	Top-down	Bottom-up
CO ₂ (Pg C/yr)	Net land CO ₂ sink	-0.4±1.1	-0.3±1.1	-0.9±1.0	-0.6±1.1	-1.6±0.9	-1.5±1.2
	1) Natural wetland	125.3±43.5	168.8±31.1	112.5±6.0	154.5±36.0	131.3±24.8	162.8±40.1
	2) Soil sinks	-15.8±6.4	-19.7±14.3	-20.3±0.0	-21.5±14.3	-24.0±6.0	-22.6±14.3
	3) Termite, Wild animal & Others	27.0±0.4	19.5±3.8	24.0±5.3	19.5±3.8	32.3±10.5	19.5±3.8
	Natural*	136.5±44.0	168.6±34.5	116.3±8.0	152.5±38.9	138.8±27.6	159.6±42.8
	4) Biomass burning	34.5±2.3	16.3±5.7	28.5±3.6	19.1±7.9	17.3±7.9	14.8±5.4
CH ₄ (Tg C/yr)	5) Rice cultivation		45.4±16.8	86.3±21.0	26.3±5.6	33.0±2.0	28.9±7.6
	6) Manure management		7.8±0.2		7.9±0.1		8.0±0.3
	7) Ruminant		64.8±2.2		66.0±0.9		70.0±3.3
	8) Landfill and Waste		33.6±2.3		39.5±2.0		44.7±3.3
	Agriculture & Waste*	156.0±12.4	151.6±17.1	179.3±45.4	139.7±6.0	168.8±26.4	151.6±9.0
	Net CH ₄ flux	327.0±45.7	336.5±38.7	324.0±46.6	311.3±39.5	324.8±38.6	325.9±43.3
	Pre-industrial CH ₄ emission			125.4±14.4			
	Human-induced CH ₄ flux	201.6±48.1	211.0±41.3	199.6±48.8	185.8±42.1	199.4±41.2	200.5±45.7
	1) Natural soil		7.9±1.3	6.6±0.5	8.2±1.3	7.5±0.4	8.4±0.9
	2) Biomass burning		0.7±0.1	0.7±0.1	0.7±0.1	0.6±0.1	0.6±0.2
	3) Agricultural soil		2.6±0.3		3.3±0.2		4.0±0.3
	4) Manure management		0.2±0.0		0.2±0.0		0.3±0.0
N ₂ O (Tg N/yr)	5) Indirect emission		0.5±0.1		0.9±0.1		0.7±0.1
	6) Human Sewage		0.2±0.6		0.2±0.0		0.3±0.0
	Agriculture & Waste *		4.7±4.2	4.1±0.6	4.6±0.2	4.4±0.6	5.5±0.7
	Net N ₂ O flux		14.0±4.3	11.3±0.8	14.3±0.9	12.6±0.7	15.2±1.0
	Pre-industrial N ₂ O emission			7.4±1.3			
	Human-induced N ₂ O flux		6.6±4.5	3.9±1.5	6.9±1.6	5.2±1.5	7.8±1.6

Estimates are derived from top-down and bottom-up approaches. The complete set of data used for these calculations can be found in Supplementary Table 2.

*Additional data sources are included in the calculation of GHG fluxes for this sub-total sector. Therefore, the sub-total GHG fluxes are not necessarily equal to the sum of individual sector values shown in this table.

Sensitivity of global terrestrial ecosystems to climate variability

Alistair W. R. Seddon^{1*}, Marc Macias-Fauria^{2*}, Peter R. Long³, David Benz³ & Kathy J. Willis^{1,3,4}

The identification of properties that contribute to the persistence and resilience of ecosystems despite climate change constitutes a research priority of global relevance¹. Here we present a novel, empirical approach to assess the relative sensitivity of ecosystems to climate variability, one property of resilience that builds on theoretical modelling work recognizing that systems closer to critical thresholds respond more sensitively to external perturbations². We develop a new metric, the vegetation sensitivity index, that identifies areas sensitive to climate variability over the past 14 years. The metric uses time series data derived from the moderate-resolution imaging spectroradiometer (MODIS) enhanced vegetation index³, and three climatic variables that drive vegetation productivity⁴ (air temperature, water availability and cloud cover). Underlying the analysis is an autoregressive modelling approach used to identify climate drivers of vegetation productivity on monthly timescales, in addition to regions with memory effects and reduced response rates to external forcing⁵. We find ecologically sensitive regions with amplified responses to climate variability in the Arctic tundra, parts of the boreal forest belt, the tropical rainforest, alpine regions worldwide, steppe and prairie regions of central Asia and North and South America, the Caatinga deciduous forest in eastern South America, and eastern areas of Australia. Our study provides a quantitative methodology for assessing the relative response rate of ecosystems—be they natural or with a strong anthropogenic signature—to environmental variability, which is the first step towards addressing why some regions appear to be more sensitive than others, and what impact this has on the resilience of ecosystem service provision and human well-being.

The rate and scale of projected climate changes in the 21st century are likely to have profound impacts on the functioning of Earth's ecosystems⁶. Much current understanding of how biodiversity will respond to climate change is based on responses to changes in mean climate state⁷. However, climate variability, and the related increases in extreme events in a warmer world⁸, has a strong influence on both the structuring and functioning of ecosystems^{9–11}. Given the importance of identifying ecologically sensitive areas for ecosystem service provision and poverty alleviation¹, a key knowledge gap exists in how to identify and then prioritize those regions that are most sensitive to climatic variability.

Ecosystem response to variability in external forcing is a key component of resilience. Theory indicates that systems with lower resilience (that is, those with a high probability of crossing a threshold to an alternative state¹²) experience amplified responses to disturbance and are more sensitive to environmental perturbations². In addition, slower responses (identified through increased autocorrelation) may be evidence of reduced recovery rates in systems approaching critical transitions¹³. Therefore, identification of areas with high ecological sensitivity or reduced recovery rates is an important step in recognizing regions of pending ecological change. In the past decade there has been an increase in the availability of satellite data measuring climate and

other ecologically relevant variables¹⁴. These data offer opportunities to characterize ecosystem sensitivity, potentially a key component of resilience, at a global scale and at high spatial resolution.

We present a novel method to identify ecosystem sensitivity to short-term climate variability and regions of amplified vegetation response (see Methods and Extended Data Fig. 1). We develop a new metric, the vegetation sensitivity index (VSI), which independently compares the relative variance of vegetation productivity (enhanced vegetation index, EVI)³ with that of three ecologically important MODIS-derived climate variables⁴ (air temperature¹⁵, water availability¹⁶ and cloud-cover)¹⁷ for each 5 km grid square for the months in which EVI and climate are found to be related. Climate–vegetation-productivity relationships are determined using an AR1 multiple linear regression approach, which uses the three climate variables and one-month-lagged vegetation anomalies (see Methods) to identify areas with strong vegetation coupling to climate anomalies (Extended Data Fig. 2). The coefficient from the one-month-lagged vegetation-productivity anomalies can be used to identify regions with memory effects, highlighting the importance of past ecosystem conditions in these regions⁵ (Extended Data Fig. 3). Our global VSI then results from aggregating the EVI sensitivities to each climate variable, weighted by the coefficients from the linear regression modelling (see Methods and Extended Data Fig. 2).

Our analysis provides three key insights into the patterns and drivers of ecological sensitivity and response to climate forcing at a global scale. First, we identify areas exhibiting amplified responses to climate variability (Fig. 1). The Arctic tundra, parts of the boreal forest belt, the wet tropical forests of South America, western Africa, and southeast Asia/New Guinea, alpine regions worldwide, steppe and prairie regions of central Asia and North and South America, the Caatinga deciduous forest in eastern South America, and eastern areas of Australia displayed high VSI values, indicating a high sensitivity to climate variability over the past 14 years. The relative contribution of each climate variable to vegetation sensitivity can also be assessed (Fig. 2). Whereas the Caatinga biome in Brazil and the prairie and grassland regions of North America and Asia are most sensitive to variations in water availability, alpine regions (for example, the Andes) demonstrate strong sensitivity to temperature, and high-latitude tundra areas exhibit strong responses to both temperature and cloud cover variability. The high sensitivity to monthly changes in cloudiness and temperature in tropical forests is also noteworthy.

Second, we present an empirical approach to quantify climate drivers of vegetation productivity (that is, the weights related to the three climate variables derived from the AR1 linear regression, Extended Data Fig. 2, hereafter climate weights, see Methods). This represents a major advancement from previous studies which have used hypothesized ecological tolerance limits to determine the relative importance of different variables driving productivity⁴. The overall picture from our empirical analysis is remarkably similar to this previous conceptual

¹Department of Biology, University of Bergen, Allégaten 41, N-500 Bergen, Norway. ²School of Geography and the Environment, South Parks Road, University of Oxford, Oxford OX1 3QY, UK.

³Long-Term Ecology Laboratory, Biodiversity Institute, Oxford Martin School, Department of Zoology, South Parks Road, University of Oxford, Oxford OX1 3PS, UK. ⁴Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK.

*These authors contributed equally to this work.

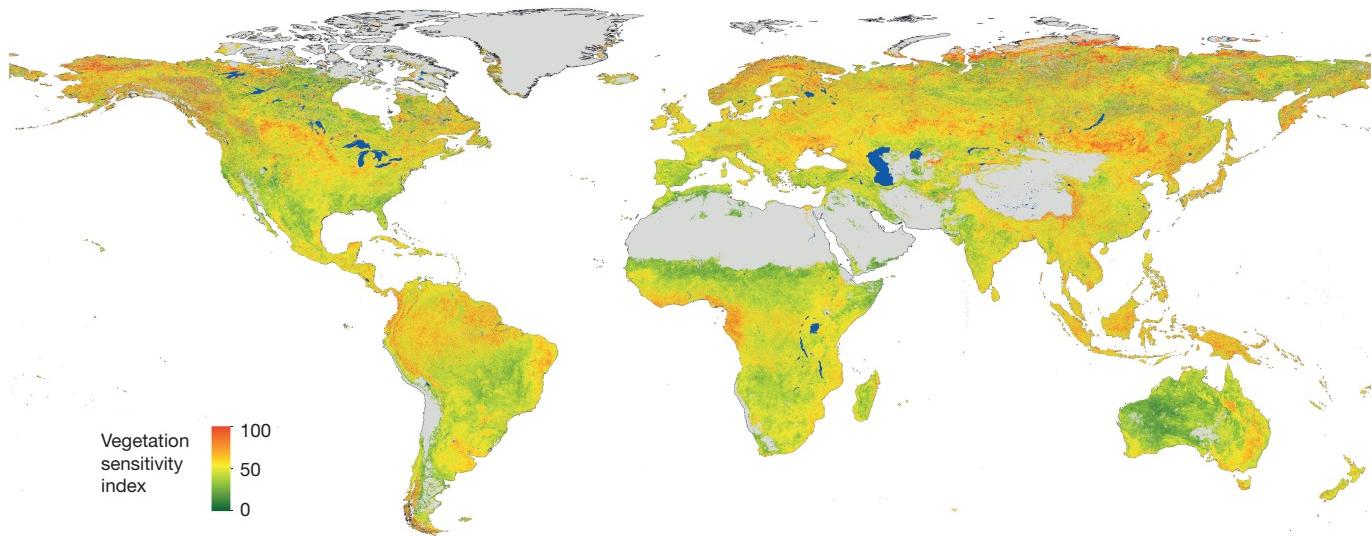


Figure 1 | Vegetation sensitivity index. Sensitivity of vegetation productivity (defined as EVI) to climate variability (based on temperature, water availability and cloudiness). The index ranges from 0 (low sensitivity, green) to 100 (high sensitivity, red). Areas with dominant barren land (mean EVI < 0.1 for all months) and permanent ice are shown grey. Wetland areas, as identified by the Global Lakes

modelling exercise⁴: prairies in mid-northern hemisphere latitudes are water limited, the high-latitudes are driven by a combination of temperature and cloudiness, and tropical forests show strong responses to cloudiness. Nevertheless, a number of key differences with this previous study are also observed. For example, central and western continental Europe exhibit stronger water limitation compared to the modelling study (as compared to temperature and radiation—a variable linked to cloudiness⁴), while water limitation was also found to be an important driver in central Africa (as compared to radiation⁴). A key question remains as to whether these differences result from modelling assumptions, or whether changing climate in the last 14 years has resulted in diverging vegetation responses in these regions.

and Wetlands Database³⁰, are mapped in blue. Pixel resolution, 5 km; period, 2000–2013. Continental outlines were modified from a shapefile using ArcGIS 10.2 software (<http://www.arcgis.com/home/item.html?id=a3cb207855b348a297ab85261743351d>). ArcGIS and ArcMap are the intellectual property of Esri and are used herein under license.

Third are the areas with high variance explained by the $t-1$ variable in the AR1 model, indicating systems where memory effects play a more important role than contemporary climate conditions in determining vegetation productivity⁵ (Extended Data Fig. 3). Overall, areas with low VSI values showed the largest memory effects (that is, high $t-1$ coefficients in our AR1 model), including the drylands of the Sahel, Australian outback, southwest USA, and the Middle East. Assessment of time series in these regions indicates that the apparent lack of response to the other climate variables occurs in two main ways: constant and largely stable low productivity conditions despite large climate variability (that is, high ecological resilience to climatic (mostly precipitation) variability, for example, Australian outback), or strong

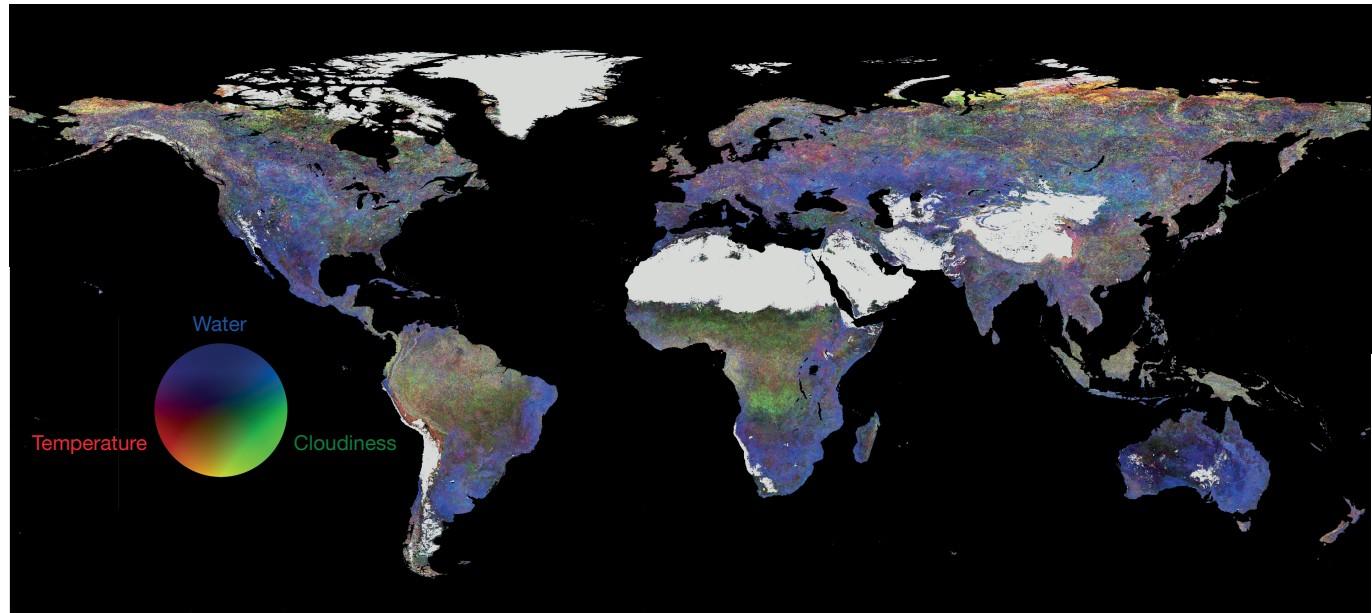


Figure 2 | RGB composite of vegetation sensitivity index. Global contribution of three climate variables to the vegetation sensitivity index (temperature, red; water availability, blue; and cloudiness, green). Pixel resolution, 5 km; period, 2000–2013. Areas with dominant barren land (mean EVI < 0.1 for all months) and permanent ice are shown grey.

cyclical variability with periods of very low and stable EVI (for example, Sahel; Extended Data Fig. 4). This contrasts to water-limited areas with higher mean EVI (for example, prairies), where strong seasonal variability is observed (Extended Data Fig. 4). Since the importance of 12-month-lagged responses in dryland regions has been previously identified¹⁸, we also tested whether model performance improved using lags of up to one year (not shown). However, we found that a one-month lag provided the best explanatory power for vegetation responses to variability on these timescales. We also found that the strength of the $t-1$ coefficient increases with decreasing levels of total annual precipitation, while there was a small positive effect on the magnitude of the climate weight related to water availability as total annual precipitation increased (Extended Data Fig. 5). These results probably indicate the importance of lagged responses to precipitation input as a result of processes related to soil-water recharge in arid regions¹⁹.

These empirically determined patterns agree with the results of multiple studies with regards to understanding current vegetation responses to climate change. Arctic and boreal regions have experienced the most rapid rates of warming in the past 30 years²⁰ and there is ample evidence on enhanced shrub growth in the tundra as a response to warming temperatures^{21,22}. We also observe similar patterns in alpine and mountainous ecosystems, adding to the increasing evidence that such areas are responding rapidly to climate change⁶. Our analysis also reveals high sensitivity to a combination of cloudiness and temperature variability in the tropical rainforest regions, particularly in the Amazon and southeast Asia (Fig. 2). Although the extent to which tropical ecosystems are currently operating at their thermal limits remains uncertain, a number of studies have found decreases in tropical forest growth rates and productivity in response to warming²³, potentially the result of reductions of leaf gas exchange under warmer temperatures²⁴. Such findings may have implications for the future of tropical forests since they are projected to experience temperature ranges beyond any current analogues²⁵. The high sensitivity to monthly changes in insolation and temperature in tropical forests observed in this study may be operating at different timescales to potential precipitation thresholds that have been identified in tropical forests²⁶. By contrast, the enhanced sensitivity to water availability in the Caatinga region of northeast Brazil agrees with studies which indicate strong coupling of vegetation cover and phenology to ENSO-related precipitation change²⁷. One potential explanation is that the high phenotypic plasticity of leaf senescence and green-up results in large amplitudes in the EVI response to drought variability. Understanding the traits that result in sensitivity differences worldwide is a key research priority.

We identified regions with high rates of response to climate variability globally and at high spatial and temporal resolutions. These properties have been linked to systems approaching ecological tipping points². However, whereas the existence of critical ecological thresholds has been suggested for a number of regions with high VSI values, such as the Arctic tundra, the boreal forest, and the wet tropical forests²⁶, some high VSI areas (for example, the steppe and prairies or the Caatinga) have not been reported to exhibit threshold-type responses at global scales²⁶. As presented, VSI is an empirically calculated state variable of ecological sensitivity for the last 14 years. As longer records of remotely sensed global vegetation and climate become available in the future, VSI offers the opportunity to identify areas showing increasing or decreasing trends in ecological sensitivity, with possible implications for identifying critical thresholds. Finally, since there is little overlap between areas demonstrating strong memory effects and those with high VSI, a question remains as to what fundamental properties underlie the difference behind fast-responding and slow-responding systems.

Identification of large-scale metrics to quantify ecological responses to climate change remains a vital strategy for global ecosystem assessment. This work builds on previous studies identifying properties that represent components of ecological resilience using satellite data^{5,28,29}. Our novel approach provides empirical baseline measurements on a

key component of ecosystem resilience, that is, the relative response of vegetation in comparison to environmental perturbations over time, as well as the climatic drivers of change across landscapes globally. The next challenge is to understand the underlying causes and ecological processes that lead to these patterns. It is also critical to determine whether these patterns represent long-lasting characteristics of the ecosystems/habitats, apparent over decades to millennia, or else more transient responses able to change spatially over short time scales, and to develop tools and technologies for modelling and predicting future trends.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 June 2015; accepted 12 January 2016.

Published online 17 February 2016.

- Convention on Biological Diversity, Aichi Biodiversity Targets. <http://www.cbd.int/sp/targets/default.shtml>
- Scheffer, M. et al. Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
- Solano, R., Didan, K., Jacobson, A. & Huete, A. MODIS vegetation index C5 user's guide (MOD13 Series), 1–42 http://vip.arizona.edu/documents/MODIS/MODIS_VI_UsersGuide_01_2012.pdf (2010).
- Nemani, R. R. et al. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
- De Keersmaecker, W. et al. A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover. *Glob. Ecol. Biogeogr.* **24**, 539–548 (2015).
- Garcia, R. A., Cabeza, M., Rahbek, C. & Araujo, M. B. Multiple dimensions of climate change and their implications for biodiversity. *Science* **344**, 1247579 (2014).
- Thomas, C. D. et al. Extinction risk from climate change. *Nature* **427**, 145–148 (2004).
- Kharin, V. V., Zwiers, F. W., Zhang, X. & Hegerl, G. C. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.* **20**, 1419–1444 (2007).
- Holmgren, M., Hirota, M., Van Nes, E. H. & Scheffer, M. Effects of interannual climate variability on tropical tree cover. *Nature Clim. Change* **3**, 755–758 (2013).
- Pederson, N. et al. The legacy of episodic climatic events in shaping temperate, broadleaf forests. *Ecol. Monogr.* **84**, 599–620 (2014).
- Doughty, C. E. et al. Drought impact on forest carbon dynamics and fluxes in Amazonia. *Nature* **519**, 78–82 (2015).
- Holling, C. S. Resilience and stability of ecological systems. *Annu. Rev. Ecol. Evol. Syst.* **4**, 1–23 (1973).
- Dakos, V. et al. Slowing down as an early warning signal for abrupt climate change. *Proc. Natl Acad. Sci. USA* **105**, 14308–14312 (2008).
- Kerr, J. T. & Ostrovsky, M. From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* **18**, 299–305 (2003).
- Seaman, S. W., Borbas, E. E., Li, J., Menzel, W. P. & Gumley, L. E. MODIS atmospheric profile retrieval algorithm theoretical basis document, version 6. http://modis-atmos.gsfc.nasa.gov/_docs/MOD07MYD07ATBDC005.pdf (2006).
- Mu, Q., Zhao, M. & Running, S. W. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.* **115**, 1781–1800 (2011).
- Ackerman, S. et al. Discriminating clear-sky from cloud with MODIS: algorithm theoretical basis document (MOD35), version 6.1. http://modisatmos.gsfc.nasa.gov/_docs/MOD35_ATBD_Collection6.pdf (2010).
- Sala, O. E., Gherardi, L. A., Reichmann, L., Jobbágy, E. & Peters, D. Legacies of precipitation fluctuations on primary production: theory and data synthesis. *Phil. Trans. R. Soc. Lond. B* **367**, 3135–3144 (2012).
- Richard, Y. & Poccard, I. A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in southern Africa. *Int. J. Remote Sens.* **19**, 2907–2920 (1998).
- Intergovernmental Panel on Climate Change. *Climate Change 2013: The Physical Science Basis*. (Cambridge Univ. Press, 2013).
- Macias-Fauria, M., Forbes, B. C., Zetterberg, P. & Kumpula, T. Eurasian Arctic greening reveals teleconnections and the potential for structurally novel ecosystems. *Nature Clim. Change* **2**, 613–618 (2012).
- Myers-Smith, I. H. et al. Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities. *Environ. Res. Lett.* **6**, 045509 (2011).
- Clark, D. A., Piper, S. C., Keeling, C. D. & Clark, D. B. Tropical rain forest tree growth and atmospheric carbon dynamics linked to interannual temperature variation during 1984–2000. *Proc. Natl Acad. Sci. USA* **100**, 5852–5857 (2003).
- Doughty, C. E. & Goulden, M. L. Are tropical forests near a high temperature threshold? *J. Geophys. Res.* **113**, G00B07 (2008).
- Williams, J. W., Jackson, S. T. & Kutzbach, J. E. Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl Acad. Sci. USA* **104**, 5738–5742 (2007).

26. Lenton, T. M. *et al.* Tipping elements in the Earth's climate system. *Proc. Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
27. Barbosa, H. A., Huete, A. R. & Baethgen, W. E. A 20-year study of NDVI variability over the northeast region of Brazil. *J. Arid Environ.* **67**, 288–307 (2006).
28. Harris, A., Carr, A. S. & Dash, J. Remote sensing of vegetation cover dynamics and resilience across southern Africa. *Int. J. Appl. Earth Obs. Geoinf.* **28**, 131–139 (2014).
29. Hirota, M., Holmgren, M., Van Nes, E. H. & Scheffer, M. Global resilience of tropical forest and savanna to critical transitions. *Science* **334**, 232–235 (2011).
30. Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol. (Amst.)* **296**, 1–22 (2004).

Acknowledgements This work was funded by Statoil ASA, Norway, Contract number 4501995279 (K.J.W., A.W.R.S., D.B.), and by the European Commission LIFE12 ENV/UK/000473 (K.J.W., D.B. and P.R.L.). P.R.L. was also

supported by an Oxford Martin School Fellowship. M.M.-F. was supported by a Natural Environment Research Council Independent Research Fellowship (NE/L011859/1) and A.W.R.S. was supported by a Research Council of Norway Postdoctoral Fellowship within a FRIMEDBIO project grant (FRIMEDBIO-214359) during analysis and write-up of this work.

Author Contributions All authors designed the study. D.B. and P.R.L. prepared and downloaded the remote-sensing data and A.W.R.S. and M.M.-F. carried out the data analysis. A.W.R.S., M.M.-F. and K.J.W. co-wrote the paper, with contributions from D.B. and P.R.L.

Author Information Remote sensing data are uploaded in the ORA repository (<http://www.bodleian.ox.ac.uk/ora>, DOI:10.5287/bodleian:VY2PeyGX4). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.W.R.S. (alistair.seddon@bio.uib.no).

METHODS

No statistical methods were used to predetermine sample size.

Satellite data. We derived monthly time series of four key ecosystem and climate variables from the MODIS sensor for the period February 2000 to December 2013. To obtain estimates of changes in ecosystem productivity, we used the MOD13C2 version 5 product which comprises monthly, global enhanced vegetation index (EVI) at 0.05° resolution³. EVI is a normalized ratio of reflectance bands with a practical range of 0 to 1. Higher values result from absorption in the visible red band of the electromagnetic spectrum. The index correlates strongly with chlorophyll content and photosynthetic activity³¹. In some cases where no clear-sky observations are available, the MOD13C2 version 5 product replaces no-data values with climatological monthly means, so we removed these values where appropriate.

We used the MOD07_L2 Atmospheric Profile product as a measure of air temperature at the same spatial resolution¹⁵. Five-minute swaths of retrieved temperature profile were projected to geographic coordinates. Pixels from the highest available pressure level, corresponding to the temperature nearest the Earth's surface, were selected in each swath. Swaths were then mean-mosaicked into global daily images, and daily images were mean-composited to monthly images to provide global time series of temperature at 0.05° resolution.

No direct estimates of incoming radiation are available from the MODIS sensor. Therefore, we developed an insolation proxy based on the MOD35_L2 Cloud Mask product¹⁷. This product provides daily records on the presence of cloudy versus cloudless skies, and we used this to make an index of the proportion of cloudy to clear-sky days in a given pixel. After conversion to geographic coordinates, five-minute swaths at 1 km resolution were re-classed as clear sky or cloudy, and these daily swaths were mean-mosaicked to global coverage, mean-composited from daily to monthly, and mean aggregated from 1 km to 0.05°. An example output from June 2005 is provided in Extended Data Fig. 6. Note that we observed a sampling bias in the MODIS insolation data at approximately 60 N in northern Eurasia, but this bias tends to occur in low insolation months between November and January and so does not influence the overall results.

The ratio of actual evapotranspiration to potential evapotranspiration (AET/PET) was used as an indicator of water availability. A value close to 1 indicates sufficient water supply to the plant, since all incoming photosynthetically active solar radiation is being used for photosynthesis. Monthly, 0.05° AET/PET was calculated from the MOD16 Global Evapotranspiration product, which estimates AET and PET through the Penman–Monteith equation^{16,32}.

Climatic drivers of vegetation productivity. To estimate the relative importance of the three climate variables driving monthly changes in productivity, all time series were transformed to z-score anomalies using monthly climatology means and standard deviations. Any month with a mean EVI below 0.1 was removed from the analysis to reduce the potential impact of noisy data at low EVI values, which are attributed to areas with extremely sparse or nonexistent vegetation cover. We also removed months with a mean monthly temperature of less than 0 °C. We then used a multiple regression approach to test for linear relationships with climate. We included the one-month-lagged EVI monthly anomalies as a fourth variable in this regression to investigate the potential influence of memory effects driving vegetation productivity⁵ (Extended Data Figs 1–3). To remove any impact of co-linearity between the three climate predictor variables³³, we used a principal components regression (PCR) to identify the relative importance of each variable driving monthly variations of EVI in each pixel. For those principal components found to have significant relationships with climate ($P < 0.1$, Extended Data Fig. 7), we multiplied the loading scores of each variable by the PCR coefficients and summed these scores. This enabled us to estimate the relative importance of each variable in driving monthly changes in productivity. Finally, we found the mean, absolute value of the variable-transformed PCR coefficients providing an empirical approach to map the relative importance of climate on productivity globally (hereafter, climate weights). The climate weights from each variable were rescaled between 0 and 1 (using the minimum and maximum value of any of the climate coefficient values) to be used for our calculations of ecological sensitivity.

Vegetation sensitivity index. To estimate ecosystem sensitivity globally, we created seasonally de-trended time series (mean monthly values subtracted) of each variable for each pixel and for periods found to have relationship with climate and the $t-1$ variable in our monthly principal components regressions. We estimated the variance of both the climatic variables and EVI on these time series. Because we found a relationship between the variance and the mean of the different months, the residuals of a quadratic linear model fitted to the mean-variance relationship of both EVI and the climate variables for each pixel were used

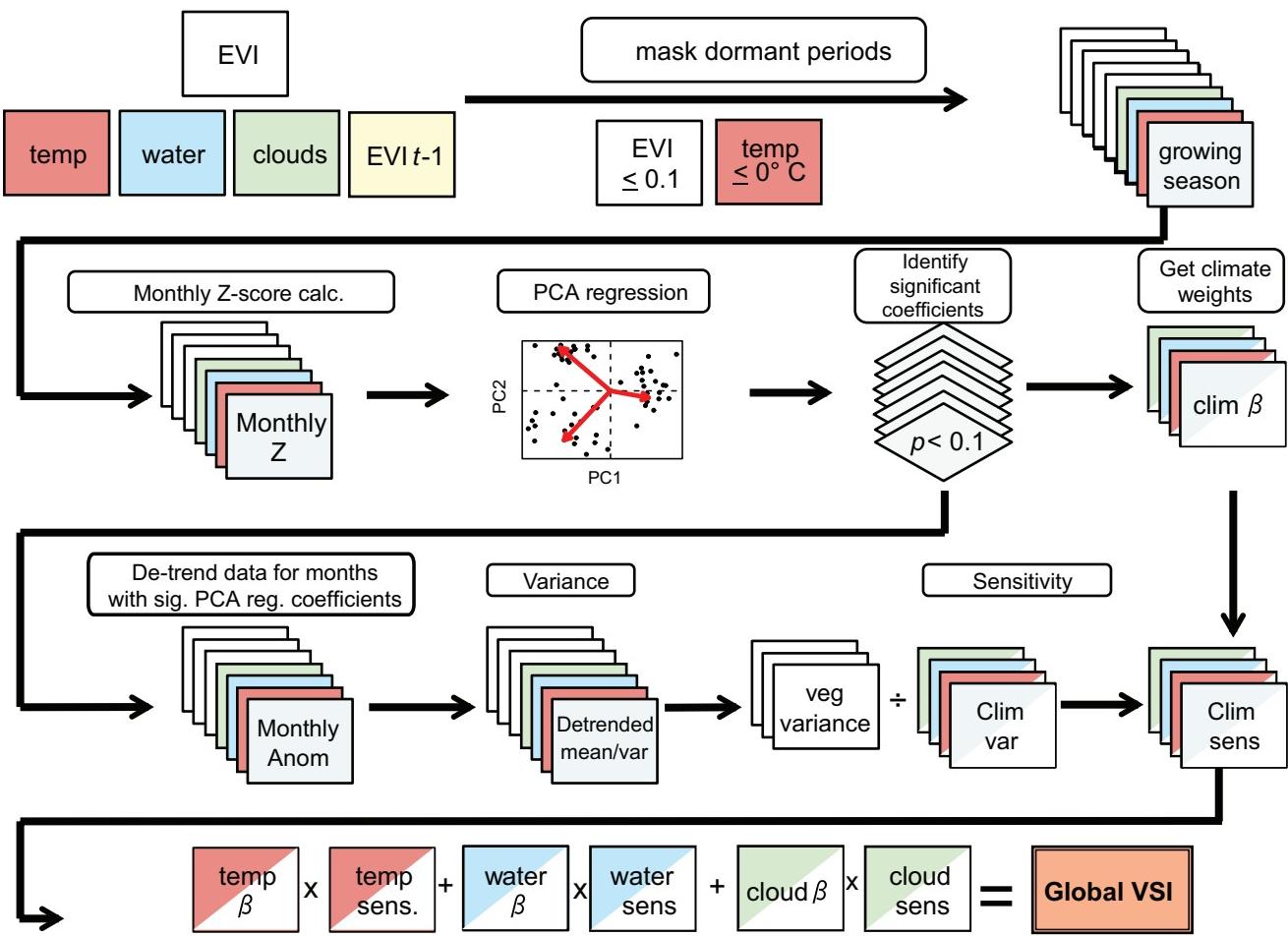
(Extended Data Fig. 8). We standardized these residuals to between 0 and 100 for each variable. Our sensitivity metrics are the \log_{10} -transformed ratios of EVI variability and each of the climate variables. Each ratio was then weighted according to the importance of the climate variable to EVI variability by multiplying it by the value of the regression coefficient (climate weights). Finally, we summed the sensitivity scores for each of our variables to identify areas of enhanced variability for the period of study (Fig. 1). All data analyses were carried out using the R project for statistical computing³⁴, using the raster³⁵, nlme³⁶, gstat³⁷, rgdal³⁸ and gtools³⁹ packages. Image processing was also carried out using Python 2.7, ArcGIS 10.2, Idrisi Selva, and the HDF-EOS to GeoTIFF Conversion Tool.

Uncertainty layers. We provide a series of maps assessing uncertainty both in the EVI measurements and in the algorithm used. In order to assess whether noise resulting from cloudy observations may be a concern to interpretations in tropical forest locations, we computed a map of the average standard error of the mean EVI score calculated for each month, which is a useful metric for identifying areas of high uncertainty in the vegetation time series (Extended Data Fig. 9). This is based on the standard deviation and number of valid EVI observations, both of which can be obtained within the metadata of the MODIS product. The highest standard errors are observed in areas with periodic presence of water on the surface (for example, Amazon river, wetlands), which is interpreted as large differences within the EVI observations and within a given month as a function of rapid, intra-month changes in the presence of surface water. Moderately high standard errors are observed in areas with more cloud cover, including parts of the wet tropical forests, the northwest coasts of Europe and North America, and some mountain ranges such as the Alps, the Pyrenees, or the Canadian Rocky Mountains. The absolute values of standard errors are not high and do not compromise the interpretation of results and their robustness: monthly EVI means for all pixels were computed from at least 25 observations on average (except for small areas in western Ecuador and Colombia, Borneo and Papua, which were based on at least 15 observations per month on average), and the monthly mean EVI standard deviation for over 90% of Earth was smaller than 0.08 (for EVI values ranging from 0 to 1).

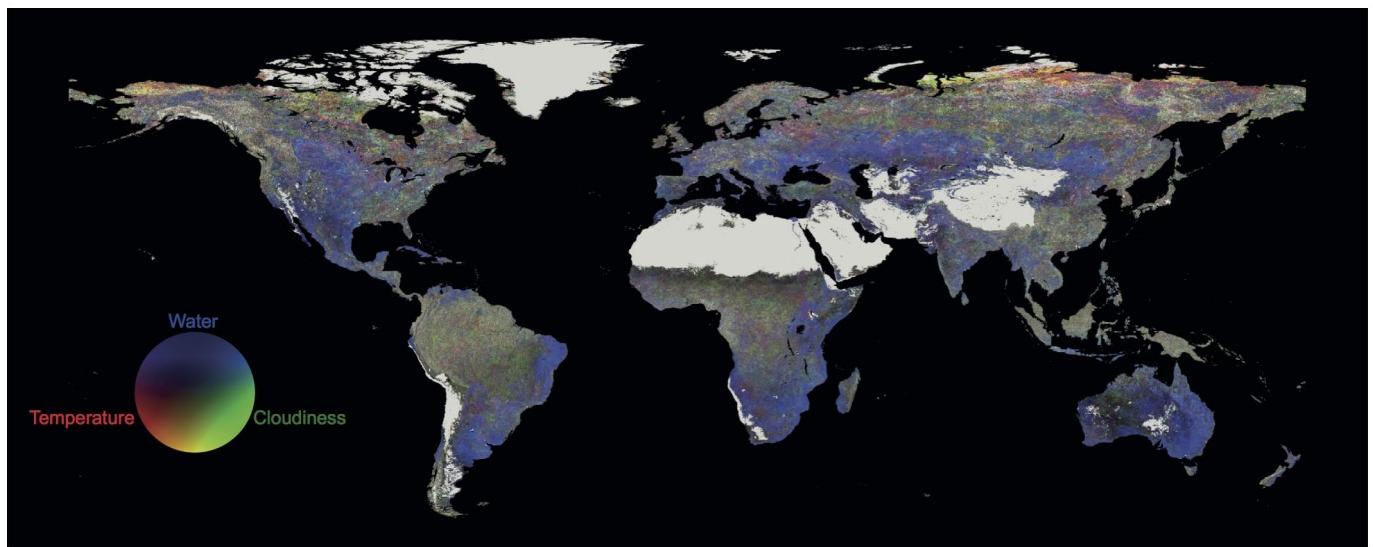
In order to assess uncertainty in our results further, we also computed confidence interval maps for every variable implemented in the regression between EVI and climate (Extended Data Fig. 10a–d). These maps were calculated by finding the upper and lower confidence intervals in the PCA regression, before transforming them back to the scale of the original climate variables using the PCA weights. We then scaled these confidence intervals by the original variables to determine uncertainty in the regression coefficients as compared to the size of the coefficients (resulting in normalized confidence interval amplitudes (NCIA)). Here, a value of 2 corresponds to a total uncertainty twice as big as the coefficient value. This analysis indicates that for all variables, NCIA is lowest where the coefficients are highest, and that the absolute NCIA values are well within acceptable levels.

Code availability. All R and MATLAB code is available for download alongside the raw data files in the ORA repository <http://www.bodleian.ox.ac.uk/ora>, DOI:10.5287/bodleian:VY2PeyGX4.

31. Huete, A. et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **83**, 195–213 (2002).
32. Cleugh, H. A., Leuning, R., Mu, Q. & Running, S. W. Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sens. Environ.* **106**, 285–304 (2007).
33. Zuur, A. F., Ieno, E. N. & Smith, G. M. *Analyzing Ecological Data*. (Springer, 2007).
34. R Core Team. R: A language and environment for statistical computing. <http://www.R-project.org> (2015).
35. Hijmans, R. J. raster: geographic data analysis and modeling. R package version 2.4-20. <http://CRAN.R-project.org/package=raster> (2015).
36. Pinheiro, J., Bates, D., DeBroy, S., Sarkar, D. & Team, A. T. R. D. C. nlme: linear and nonlinear mixed effects models. R package version 3.1-122. <http://CRAN.R-project.org/package=nlme> (2013).
37. Pebesma, E. J. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* **30**, 683–691 (2004).
38. Bivand, R., Keitt, T. & Rowlingson, B. rgdal: bindings for the geospatial data abstraction library. R package version 0.9-3. <http://CRAN.R-project.org/package=rgdal> (2015).
39. Warnes, G. R., Bolker, B. & Lumley, T. gtools: various R programming tools. R package version 3.5.0. <http://CRAN.R-project.org/package=gtools> (2015).
40. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
41. Pope, N. corHaversine function. <http://stackoverflow.com/questions/18857443/specifying-a-correlation-structure-for-a-linear-mixed-model-using-the-ramps-pack> (2013).

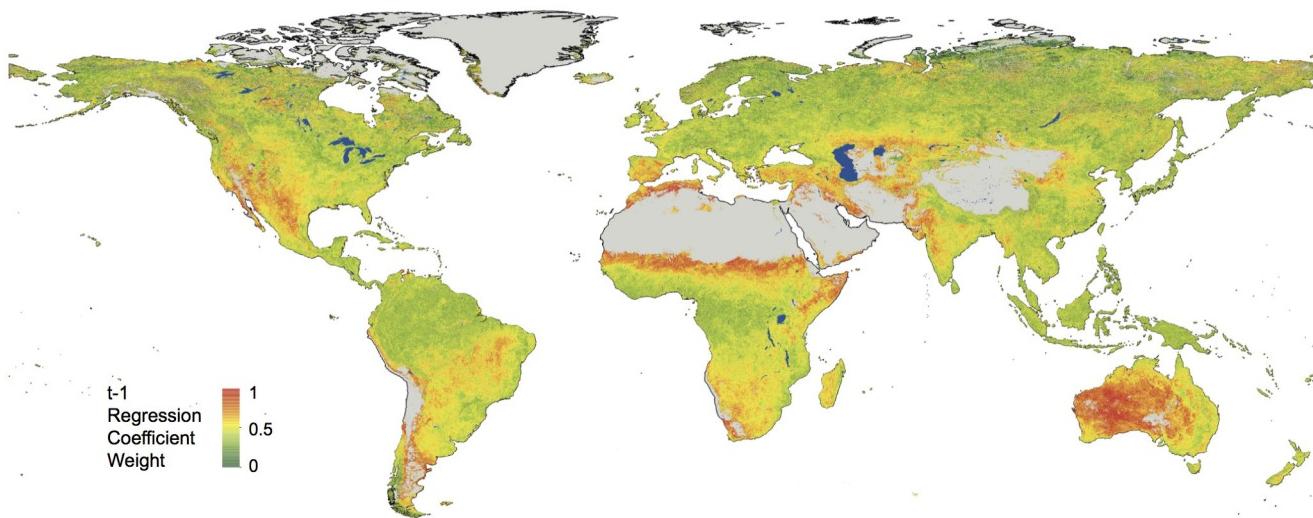


Extended Data Figure 1 | Study Design. Flow chart of the algorithm used to estimate the vegetation sensitivity index.



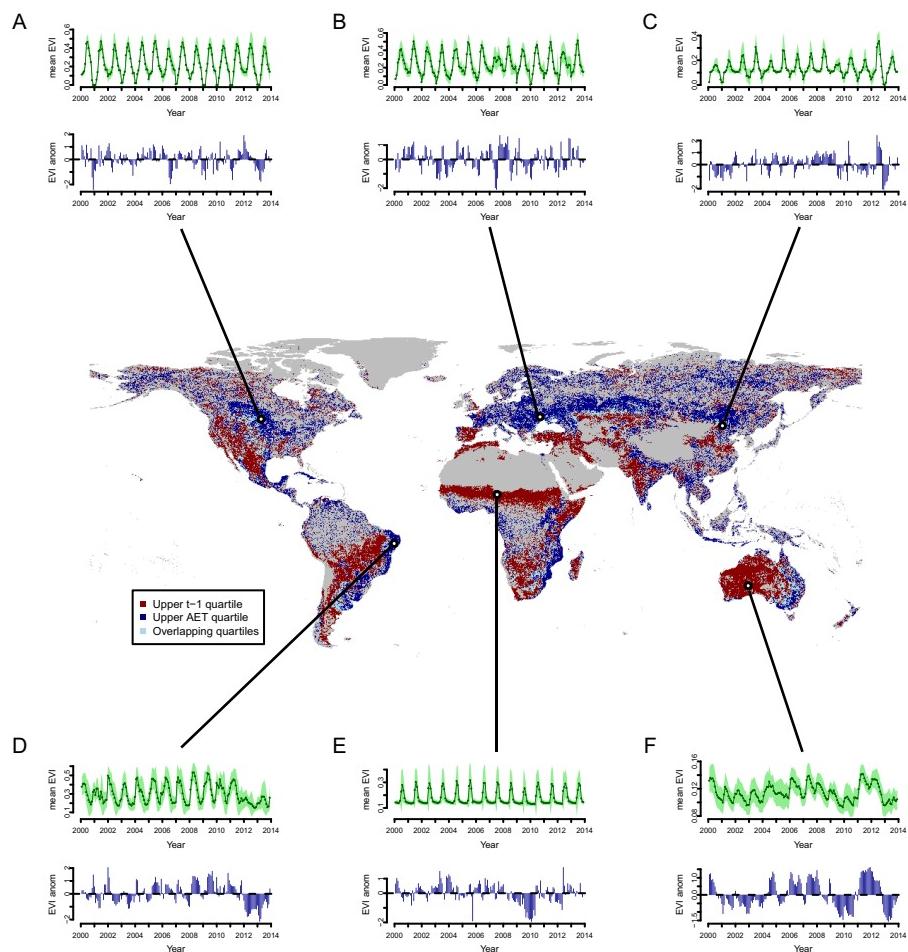
Extended Data Figure 2 | RGB composite of climate weights. RGB composite global map of the mean climate coefficient weights from monthly multiple regressions between vegetation productivity (defined as EVI), vegetation productivity at $t-1$ and three climate variables

(temperature, red; water availability, blue; and cloudiness, green). Areas with dominant barren land ($\text{mean EVI} < 0.1$ for all months) and permanent ice are shown grey. Pixel resolution, 5 km; period, 2000–2013.


Extended Data Figure 3 | Global map of the $t-1$ coefficient.

Global map of $t-1$ (AR1) coefficient weight from a monthly multiple regressions between vegetation productivity (defined as EVI), vegetation productivity at $t-1$ and the three climate variables. Areas with dominant barren land (mean EVI < 0.1 for all months) and permanent ice are shown grey. Wetland areas, as identified by the Global Lakes

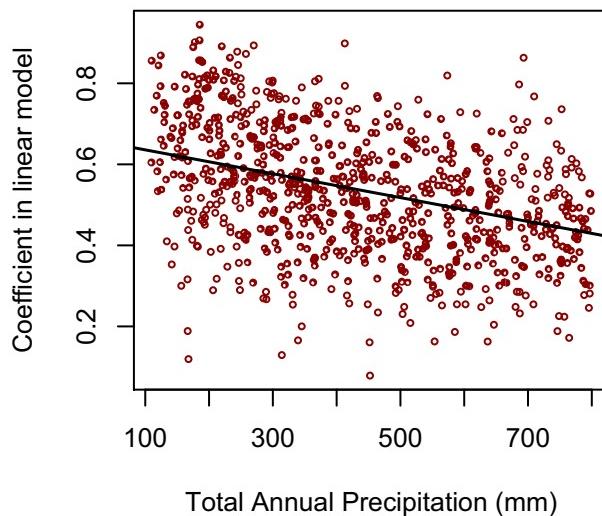
and Wetlands Database³⁰, are mapped in blue. Pixel resolution, 5 km; period, 2000–2013. Continental outlines were modified from a shapefile using ArcGIS 10.2 software (<http://www.arcgis.com/home/item.html?id=a3cb207855b348a297ab85261743351d>). ArcGIS and ArcMap are the intellectual property of Esri and are used herein under license.



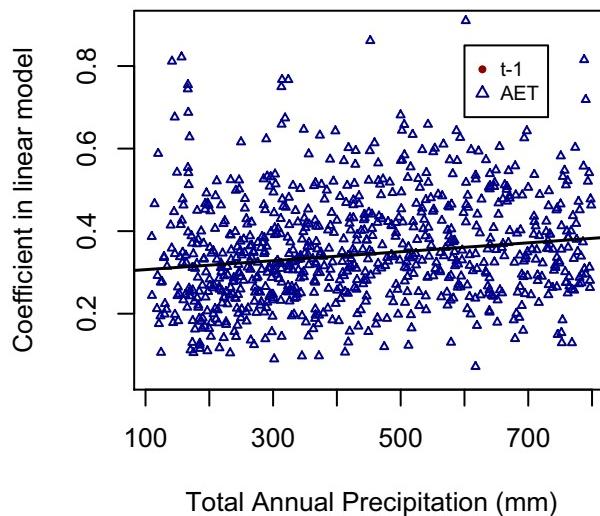
Extended Data Figure 4 | EVI variability in areas of low total annual precipitation. Time series plots of the mean EVI (green) and mean EVI monthly anomalies (blue) for six different dryland/water-limited regions across the world. Time series are calculated by finding the mean monthly value for all 5-km pixels with a 1° grid cell (total pixels = 400). The light green shading in the mean EVI plots represents the upper and lower two standard deviations. **a**, North American temperate grassland

(pixel centre 99.5°W, 47.5°N). **b**, Eurasian temperate grassland (30.5°E, 48.5°N). **c**, Eurasian temperate grassland (115.5°E, 44.5°N). **d**, Caatinga forests, woodlands and scrub (37.5°W, 8.5°S). **e**, Sahel subtropical savanna and shrubland (10.5°E, 13.5°N). **f**, Australian desert (127.5°E, 27.5°N). The map in the main panel insert represents areas with $t-1$ and water limitation linear regression coefficients within the upper quartile (see Methods). Red, $t-1$; dark blue, water limitation; light blue, both).

A

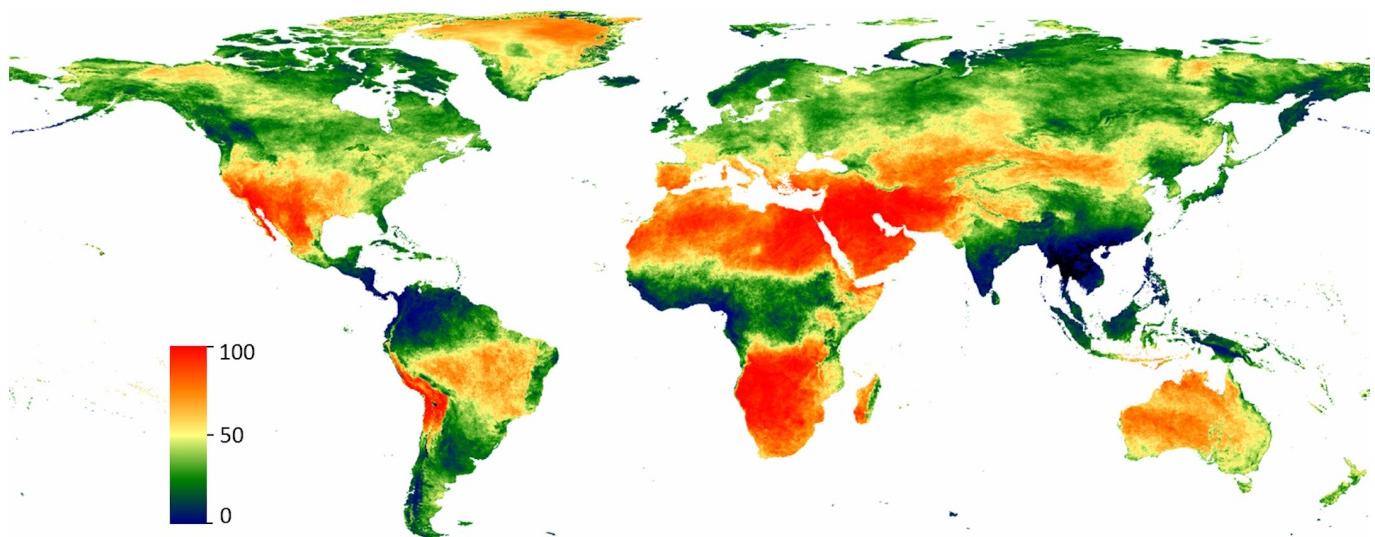


B

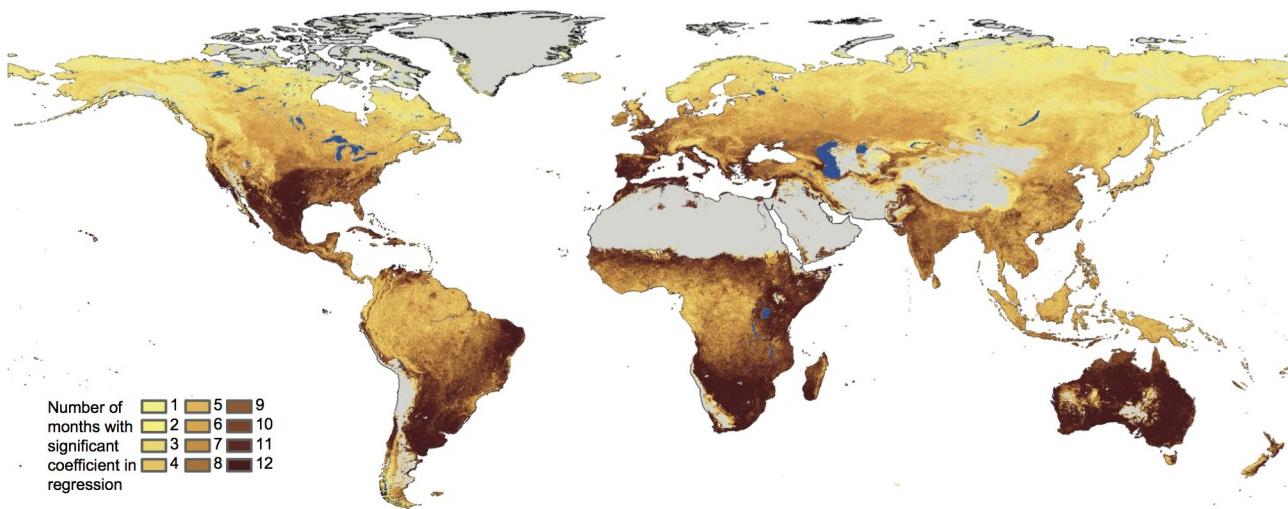


Extended Data Figure 5 | $t-1$ and water limitation against total annual precipitation. **a, b,** Plots of the $t-1$ (a) and water limitation coefficients (b) from the AR1 linear regression model (see Methods) plotted against total annual precipitation (mm) calculated as the sum of the WorldClim monthly precipitation data⁴⁰. A random subsample of 1,000 points were taken from dryland areas, defined here as having total annual precipitation between 100 – 800 mm, and between 50 N and 50 S. After removing no-data values from the random subset (that is, unresponsive pixels from the VSI calculation), the total number of samples was 795. A linear model

was fit to both data sets independently using generalized least squares in the ‘nlme’³⁶ package in R³⁴. An exponential spatial error term using geographic distance was used to account for spatial autocorrelation in the residuals in the model⁴¹. There was a negative significant effect on the size of the $t-1$ coefficient with increasing total annual precipitation (-0.0003 ± 0.00003 , significant at $P < 0.01$), with a smaller, positive effect of total annual precipitation on water availability (0.0001 ± 0.00003 , significant at $P < 0.01$).

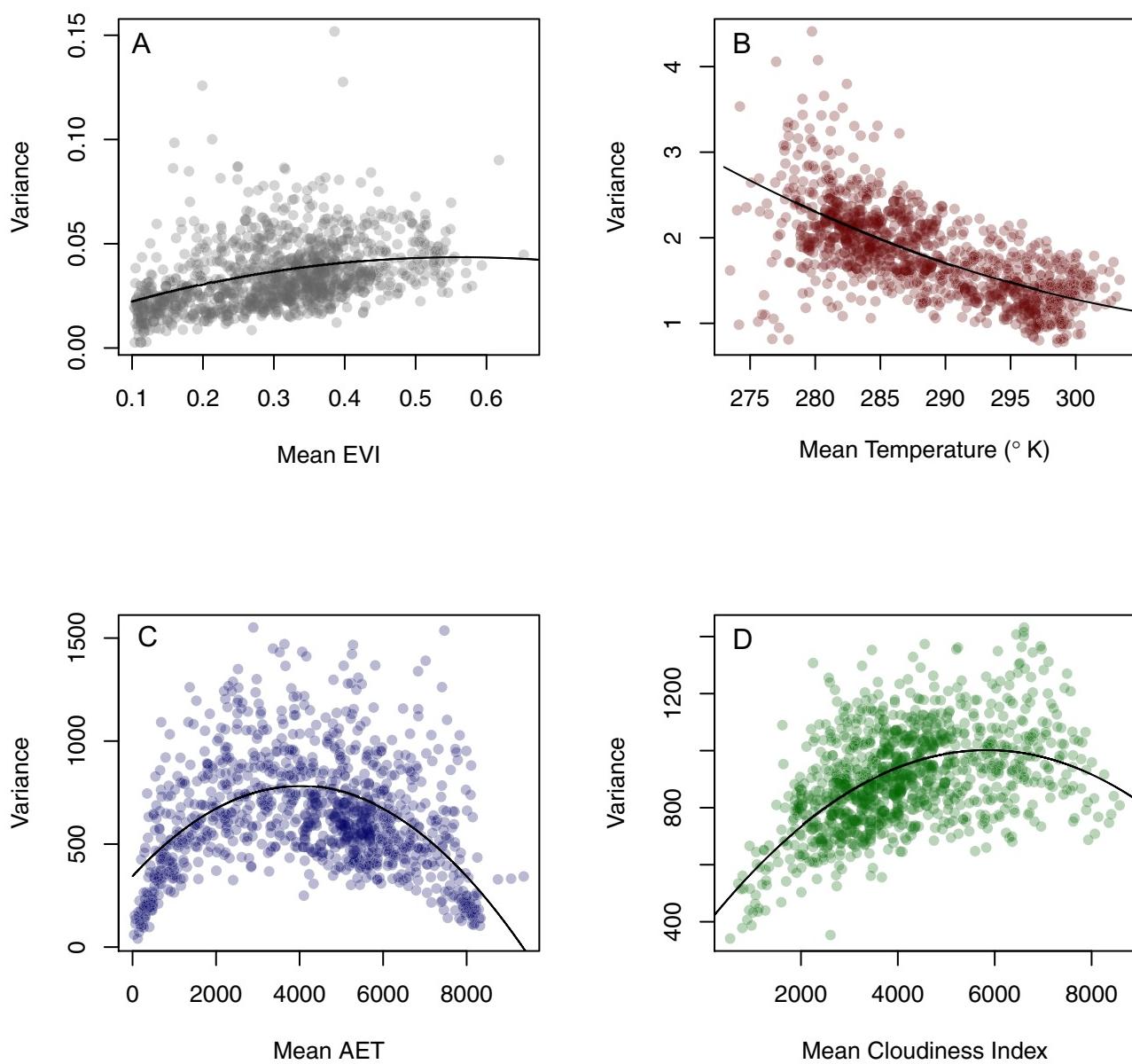


Extended Data Figure 6 | Cloudiness index. Example output of the cloudiness index derived from the MOD35_L2 Cloud Mask product for June 2005. High values indicate more cloud-free days. Note the large number of cloud-free days in dryland regions, and the large number of cloudy days in southeast Asia as a result of the seasonal monsoon. Pixel resolution, 5 km.

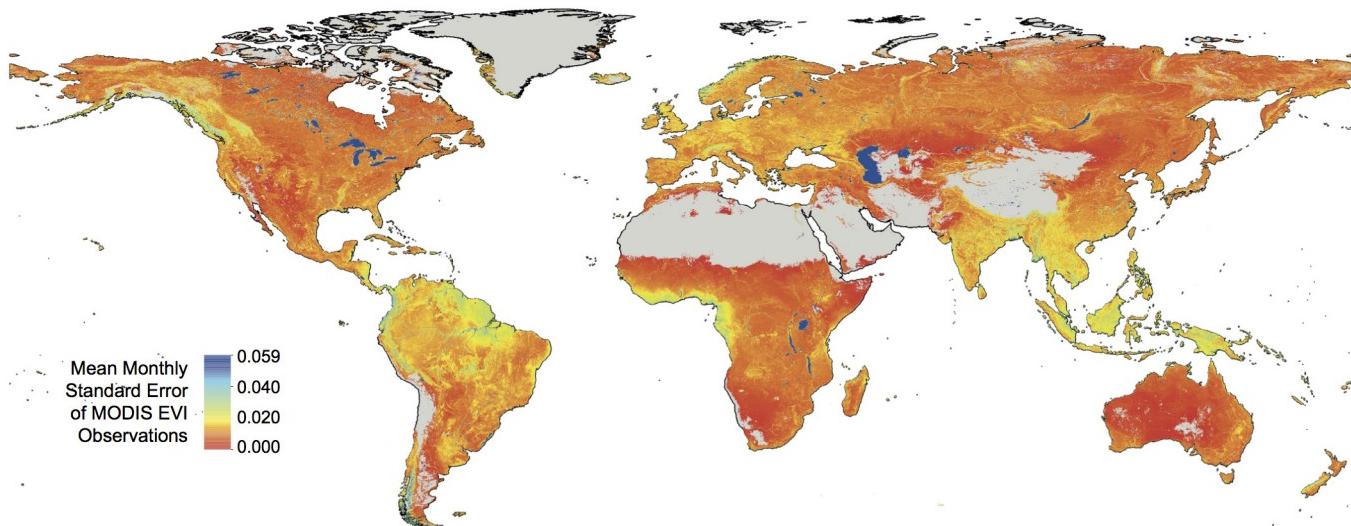


Extended Data Figure 7 | Number of months with a significant ($P < 0.1$) coefficient in the principal components regression. Number of months with a significant ($P < 0.1$) coefficient in the principal components regression between vegetation productivity (EVI), and climate (temperature, water availability, and cloud cover), and a $t-1$ vegetation variable. Areas with dominant barren land (mean EVI < 0.1 for all months) and permanent ice are shown grey. Wetland areas, as identified

by the Global Lakes and Wetlands Database³⁰, are mapped in blue. Pixel resolution, 5 km; period, 2000–2013. Continental outlines were modified from a shapefile using ArcGIS 10.2 software (<http://www.arcgis.com/home/item.html?id=a3cb207855b348a297ab85261743351d>). ArcGIS and ArcMap are the intellectual property of Esri and are used herein under license.

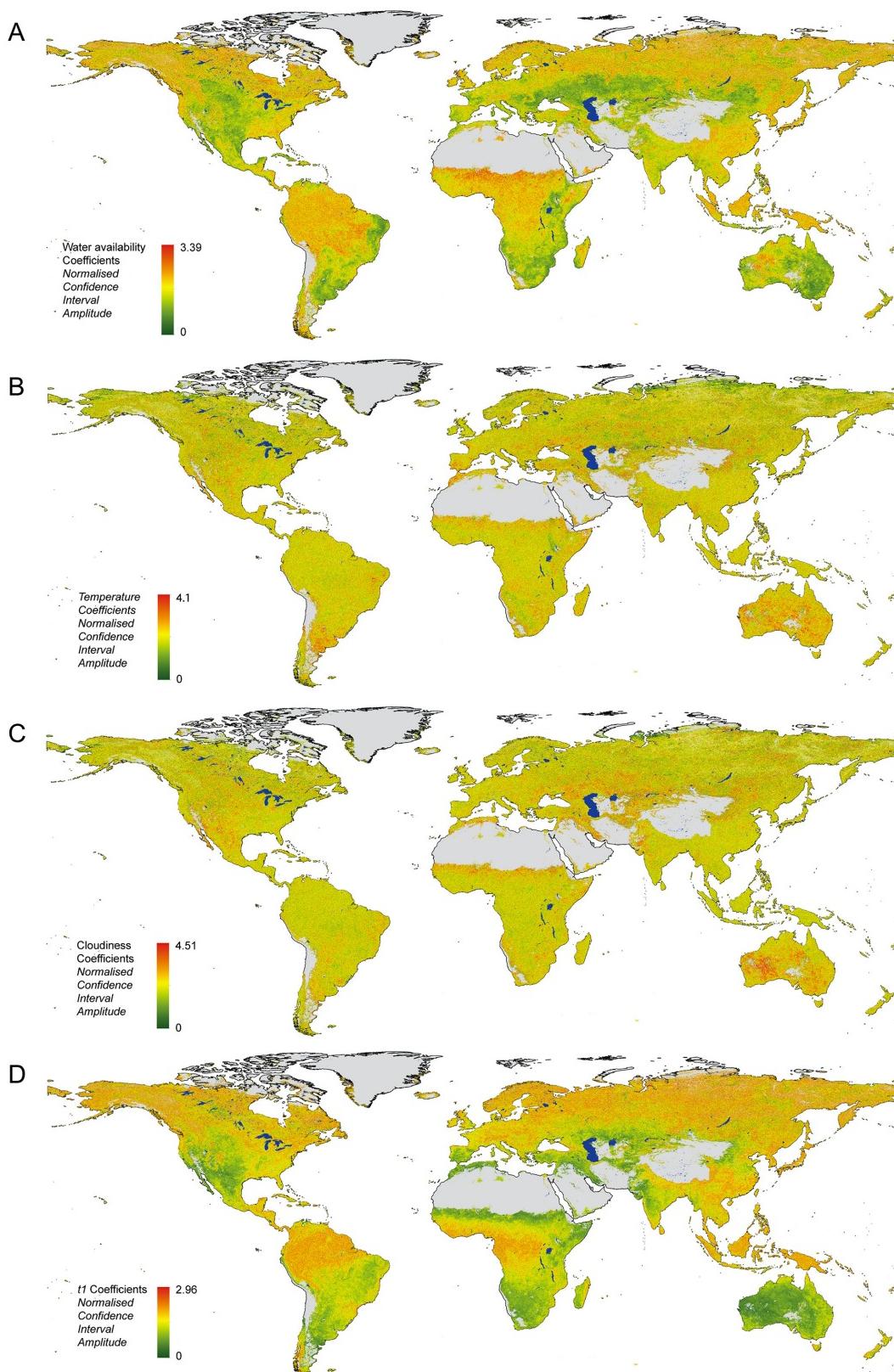


Extended Data Figure 8 | Mean–variance relationships. a–d, Plots of the mean–variance relationships for EVI (a) and the three climate variables derived from MODIS data (ground temperature (b), water availability (c) and cloud cover (d)). Owing to the large number of pixels ($7,200 \times 3,000$), these plots are made using 1,000 randomly sampled points from across the Earth surface for clarity.



Extended Data Figure 9 | Mean standard error of the MODIS EVI observations. Mean standard error of the MODIS EVI observations, calculated on a monthly basis over the period 2000–2013 as the standard deviation of all EVI observations per 5 km pixel divided by the square root of the number of observations. Areas with dominant barren land (mean EVI < 0.1 for all months) and permanent ice are shown grey. Wetland

areas, as identified by the Global Lakes and Wetlands Database³⁰, are mapped in blue. Continental outlines were modified from a shapefile using ArcGIS 10.2 software (<http://www.arcgis.com/home/item.html?id=a3cb207855b348a297ab85261743351d>). ArcGIS and ArcMap are the intellectual property of Esri and are used herein under license.



Extended Data Figure 10 | Normalized confidence interval amplitudes.

Normalized confidence interval amplitudes (NCIA) for the regression coefficients in the EVI versus external forcings (temperature, water availability, cloudiness) and memory effects (EVI $t-1$) regression. Larger NCIA values correspond to larger uncertainty in the coefficient estimates. Amplitudes were normalized by the mean coefficient value in each 5 km pixel (that is, a value of 2 corresponds to a total uncertainty twice as big as the coefficient value). Only significant coefficients in the original PCA regression were accounted for, and hence no coefficient crosses zero in any

pixel. Areas with dominant barren land ($\text{mean EVI} < 0.1$ for all months) and permanent ice are shown grey. Wetland areas, as identified by the Global Lakes and Wetlands Database³⁰, are mapped in blue. **a**, Water availability; **b**, temperature; **c**, cloudiness; **d**, EVI $t-1$. Continental outlines were modified from a shapefile using ArcGIS 10.2 software (<http://www.arcgis.com/home/item.html?id=a3cb207855b348a297ab85261743351d>). ArcGIS and ArcMap are the intellectual property of Esri and are used herein under license.

Sex speeds adaptation by altering the dynamics of molecular evolution

Michael J. McDonald^{1,2*}, Daniel P. Rice^{1,2*} & Michael M. Desai^{1,2,3}

Sex and recombination are pervasive throughout nature despite their substantial costs¹. Understanding the evolutionary forces that maintain these phenomena is a central challenge in biology^{2,3}. One longstanding hypothesis argues that sex is beneficial because recombination speeds adaptation⁴. Theory has proposed several distinct population genetic mechanisms that could underlie this advantage. For example, sex can promote the fixation of beneficial mutations either by alleviating interference competition (the Fisher–Muller effect)^{5,6} or by separating them from deleterious load (the ruby in the rubbish effect)^{7,8}. Previous experiments confirm that sex can increase the rate of adaptation^{9–17}, but these studies did not observe the evolutionary dynamics that drive this effect at the genomic level. Here we present the first, to our knowledge, comparison between the sequence-level dynamics of adaptation in experimental sexual and asexual *Saccharomyces cerevisiae* populations, which allows us to identify the specific mechanisms by which sex speeds adaptation. We find that sex alters the molecular signatures of evolution by changing the spectrum of mutations that fix, and confirm theoretical predictions that it does so by alleviating clonal interference. We also show that substantially deleterious mutations hitchhike to fixation in adapting asexual populations. In contrast, recombination prevents such mutations from fixing. Our results demonstrate that sex both speeds adaptation and alters its molecular signature by allowing natural selection to more efficiently sort beneficial from deleterious mutations.

The vast majority of species engage in some form of sex or genetic exchange¹. Yet the evolutionary forces that make sex widespread in nature remain incompletely understood. In principle, asexual reproduction should be more efficient: it avoids the costs of mating and allows individuals to pass all (rather than half) of their genetic material to their offspring. Extensive theoretical work has sought to understand why sex is pervasive despite these substantial costs^{2,3}.

One potential evolutionary advantage of sex is that recombination can speed adaptation⁴. Several distinct mechanisms could drive this effect. For example, recombination can relieve clonal interference, bringing together beneficial mutations that arise on different genetic backgrounds and would otherwise compete^{5,6,18–20}. Sex can also rescue beneficial mutations from deleterious backgrounds^{7,8}. Recent empirical work suggests that such interference effects are widespread in adapting asexual microbial^{21,22} and viral populations²³, and may also be common in higher eukaryotes²⁴. Thus the role of recombination in speeding adaptation may be broadly important in the evolution and maintenance of sexual reproduction.

Several laboratory evolution experiments have confirmed that sex can indeed increase the rate of adaptation^{9–17}. By analysing how the strength of this effect depends on population size^{13,14} and other parameters^{9–12}, these studies sought to quantify the relative importance of various potential advantages of sex. However, previous studies have been limited almost exclusively to phenotypic measurements. Hence

they have been unable to observe how recombination alters evolutionary dynamics at the sequence level. This has made it difficult to connect phenotypic observations of the advantages or disadvantages of sex to their underlying molecular causes.

Here we describe the first comparison of the dynamics of genome sequence evolution in sexual and asexual populations. We use experimental evolution of *S. cerevisiae* as a model system. As in earlier studies^{11,12}, we incorporate recombination by interspersing asexual mitotic growth (with mating type **a** and **α** subpopulations propagated separately) with discrete ‘sexual cycles’ of mating followed by sporulation (Methods). Sexual cycles pose a key technical challenge: it is difficult to ensure that most of the population sporulates and mates without inbreeding. To overcome this obstacle, we developed a genetic system involving two drug markers, one tightly linked to each mating locus, combined with haploid-specific and mating-type-specific nutrient markers (Extended Data Fig. 1). This enabled us to force outcrossing by selecting separately for haploid **a** and **α** cells after sporulation and for diploids after mating. We verified that leakage of mitotically dividing cells through each cycle is minimal (<0.1%), and that sexual cycles do not introduce bottlenecks compared with the effective population size (Methods and Extended Data Table 1). This system allows us to control the rate of outcrossing, and hence isolate the effects of recombination from ancillary features of the experimental protocol.

Using this approach, we evolved 6 replicate sexual populations and 12 asexual controls (each consisting of a single type **a** or **α** population). Each population was founded from a single clone and propagated at an effective population size of $\sim 10^5$ cells (Methods). We induced sex every 90 generations. During sexual cycles, we ensured that selection pressures in sexual and asexual lines were as equivalent as possible (without inducing mating or sporulation in asexuals; Methods). We verified that any differences between these treatments do not lead to differential adaptation to sexual cycles (or asexual control conditions) by measuring how sexual and asexual lines adapted to both conditions (Extended Data Fig. 2). We also confirmed that these conditions do not lead to different mutation rates (Extended Data Table 2). We note that each sexual line consists of a mating type **a** and type **α** subpopulation, while asexual lines consist of a single type **a** or **α** population, creating a potential difference in effective population size. To verify that this does not affect our conclusions, we evolved a parallel set of asexual control lines, each consisting of two type **a** subpopulations mixed at 90-generation intervals (analogous to sexual lines but without recombination). We confirmed that these lines adapt at the same rate as asexual lines consisting of a single subpopulation each (Methods; Extended Data Fig. 3).

After $\sim 1,000$ generations of adaptation, including 11 sexual cycles, we measured the fitness of multiple clones isolated from each population (Methods; note one sexual population ended at generation 900 owing to technical failures during evolution). We also measured the fitness

¹Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ²FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ³Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

*These authors contributed equally to this work.

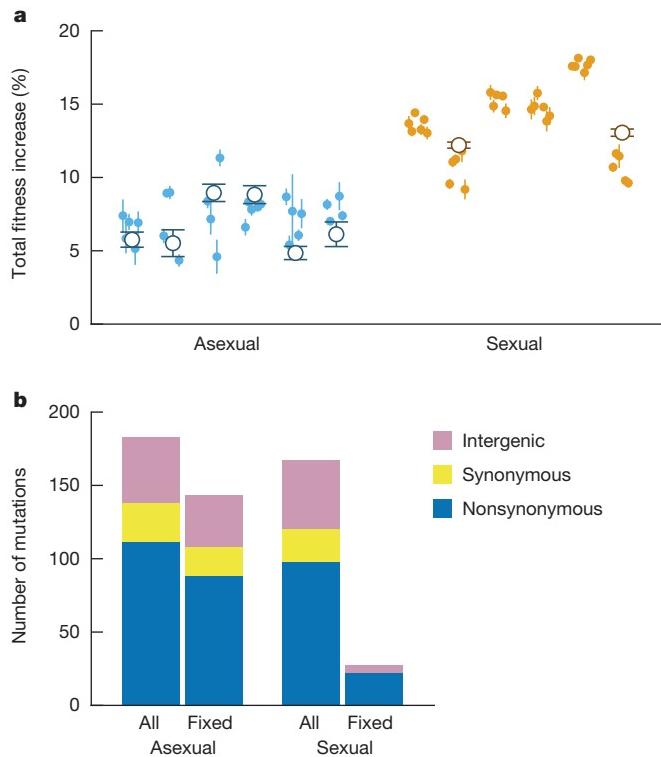


Figure 1 | The rate and molecular signatures of adaptation. **a**, Total fitness increase over $\sim 1,000$ generations of adaptation in asexual (blue) and sexual (orange) populations. Open circles represent the mean fitness of the population (for type **a** populations without frequency dependence); solid points represent the fitness of individual clones (mean of five replicate fitness assays, error bars \pm s.e.m.). **b**, Classification of observed and fixed mutations in sequenced lines.

of whole-population samples, except in four sexual populations where the spontaneous evolution of frequency-dependent interactions made population fitness undefined (we describe this frequency dependence below). Both clone (Mann–Whitney *U*-test, $P < 0.001$) and whole-population (two-sided *t*-test, $P < 0.001$) fitness data show that sexual populations adapted significantly faster than asexual controls (Fig. 1a).

To reveal the molecular mechanisms underlying faster adaptation in sexual populations, we turned to whole-genome sequencing. We sequenced whole-population samples every 90 generations in four sexual and four asexual populations. We identified segregating mutations and tracked their frequencies through time (Methods). We detected an average of 44 *de novo* mutations per population (Extended Data Table 3 and Supplementary Data 1). We emphasize that these results represent a subset of all mutations in our populations. Most importantly, we focus on single nucleotide polymorphisms (SNPs) and small indels; we cannot call certain more complex types of mutation (for example, large indels and chromosomal rearrangements) from whole-population data. To estimate the impact of these complex mutations, we sequenced eight total clones isolated from two sexual and two asexual populations, identifying no aneuploidies and only a small number (~ 2.5 per population) of duplications and deletions of at most 65 kb (half in transposable elements; Methods, Extended Data Fig. 4 and Extended Data Table 4). Since we cannot track them in whole-population data, we neglect these events in our analysis.

We find that sex alters the molecular signatures of adaptation. We observe similar proportions of synonymous, nonsynonymous, and intergenic mutations segregating in sexual and asexual lines (Fig. 1b). Consistent with earlier work²¹, in asexual populations these types of mutation are roughly equally likely to fix, conditional on reaching observable frequency (Fig. 1b and Extended Data Table 3). This indicates that natural selection cannot efficiently distinguish between

their effects. In contrast, fewer mutations fix in sexual populations, and these mutations are overwhelmingly nonsynonymous. These observations suggest that sex improves the efficiency of selection, so that only beneficial mutations fix.

To investigate how sex improves the efficiency of selection, we analysed the dynamics of adaptation. As in earlier studies^{21,22}, asexual populations exhibit signatures of hitchhiking and clonal interference (Fig. 2a–d). Groups of functionally unrelated mutations, linked within the same genetic background, change in frequency together as clonal cohorts. The outcomes of evolution are determined by competition between these cohorts. In contrast, sexual populations are not characterized by cohorts of linked mutations (Fig. 2e–h). Instead, the dynamics of each mutation is largely independent of other variation in the population. In these populations, mutations that occur on different backgrounds fix independently, while others briefly hitchhike to moderate frequencies where they persist or are eliminated from the population.

We quantified these differences in dynamics by calculating the correlations in frequency changes between mutations (Methods). This measures how linked or independent the fates of these mutations are (for example, linked mutations within clonal cohorts are strongly correlated). As expected, we find stronger correlations in asexual populations (Kolmogorov–Smirnov test, $P < 10^{-6}$; Fig. 2i). We also compared the correlations within each population to a null distribution of correlations between trajectories in different populations (Methods). Both sexual and asexual populations exhibit stronger correlations than the null expectation (Kolmogorov–Smirnov test, $P < 10^{-6}$; Fig. 2j, k), but the deviation is stronger in asexuals (Fig. 2l).

These differences in the dynamics and molecular signatures of adaptation suggest that recombination makes natural selection more efficient at fixing beneficial mutations and purging neutral or deleterious hitchhikers, as argued by earlier studies⁹. For example, in asexual populations some cohorts that initially increase in frequency are later driven to extinction (Fig. 2a–d), consistent with earlier work²¹. This indicates that adaptation in asexuals is limited by competition between cohorts that drives some beneficial mutations extinct. To analyse the efficiency of selection more directly, we measured fitness effects of individual mutations using two methods. First, we used a sequencing-based fitness assay. Specifically, we crossed an evolved clone from each sequenced population to its ancestor, generating a bulk segregant pool in which each mutation is present in many genetic backgrounds. We propagated this pool for 70 generations, sequenced at four time points, and tracked the frequency of each mutation to measure its fitness effect averaged across backgrounds (Methods). Second, we selected four genes that were mutated in both an asexual and sexual population, reconstructed each in a corresponding ancestral or evolved clone, and measured their fitness effects (Methods).

As expected, we find that each clonal cohort that fixes in an asexual population contains at least one beneficial mutation. However, we also find that significantly deleterious mutations hitchhike to fixation (Fig. 3a, c). Recent theory has argued that the fixation of strongly deleterious mutations can be common in adapting asexual populations^{25,26}. Our results provide the first direct experimental support for this hypothesis. In contrast, recombination decouples hitchhiking mutations from their initial background, and we identify no deleterious mutations that fix in sexual populations (Fig. 3b, c). The potential for sex to purge deleterious mutations in non-adapting populations has been extensively studied²⁷ (for example, in work on Muller's ratchet). Our experiments show that this effect is important even in adapting populations, confirming recent theory^{28,29}.

Our genetic reconstructions also highlight the potential importance of epistasis. For example, we identified a mutation in *MET2* that fixed in a sexual population despite being deleterious in the ancestral background. However, further reconstructions showed that this mutation is beneficial in an evolved background, an example of sign epistasis (Methods). We cannot rule out the possibility of similar epistatic effects

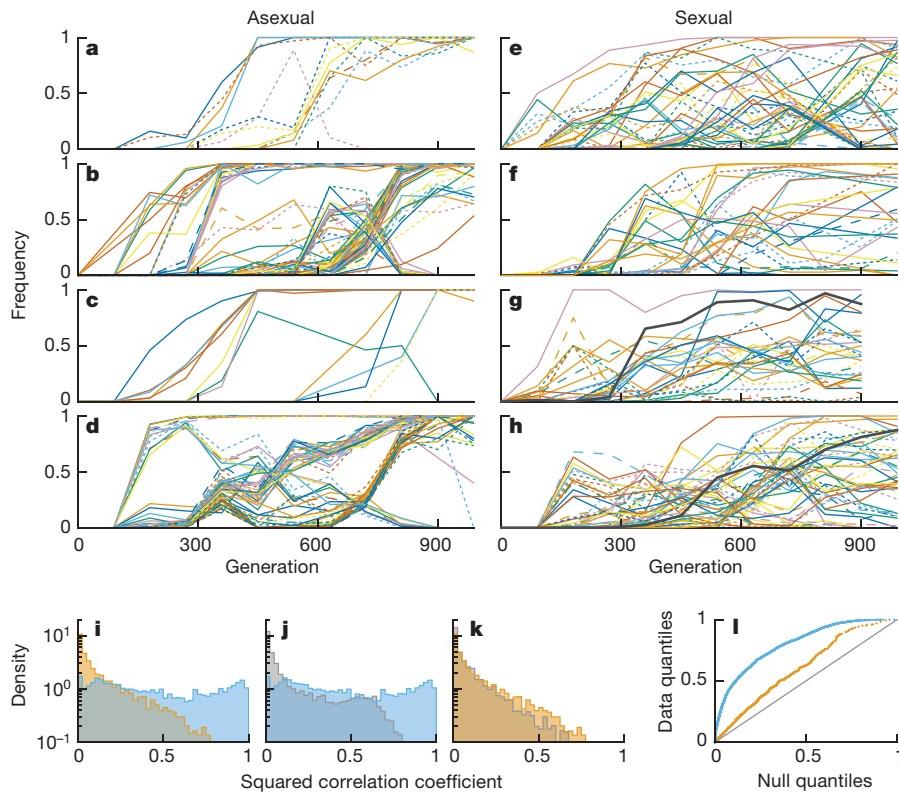


Figure 2 | Fates of spontaneously arising mutations. **a–h**, The frequencies of all identified *de novo* mutations through ~1,000 generations in four asexual populations (**a–d**) and four sexual populations (**e–h**). Solid lines are nonsynonymous mutations; dashed are synonymous; dotted are intergenic. Black trajectories represent mutations in *ERG3* subject to balancing selection. **i**, Distribution of correlations in frequency

changes among pairs of trajectories in asexual (blue) and sexual (orange) populations. **j, k**, Comparison between these correlations and an empirical null distribution (grey) in asexual (**j**) and sexual (**k**) populations. **l**, Quantile–quantile plot summarizing deviations from null expectations (grey) in asexual (blue) and sexual (orange) populations.

involving other mutations; this represents a limitation of the analysis in Fig. 3.

Four sexual populations spontaneously evolved an ‘adherent’ phenotype that stably coexists with the wild type. In earlier work³⁰, we showed that this adherent type arises owing to a loss-of-function mutation in the ergosterol pathway, which is maintained by balancing selection.

Sequencing two of these populations revealed distinct mutations in *ERG3*, which persist at intermediate frequencies (Fig. 2g, h). Despite the stable coexistence of these two phenotypes, our sequence data demonstrate that other mutations recombine between types before sweeping through the entire population. In combination with our fitness data, these results show that sex speeds adaptation despite the

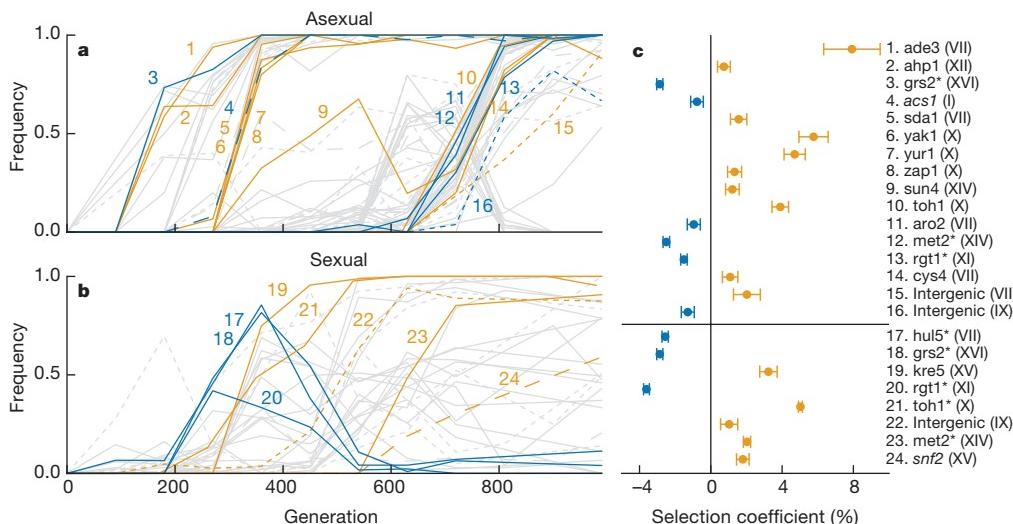


Figure 3 | Fitness effects of individual mutations. **a, b**, Mutation trajectories in an asexual (**a**) and sexual (**b**) line. Orange mutations are significantly beneficial; blue are deleterious; grey are unmeasured or consistent with neutrality. Several mutations were present in the founding population and hence omitted from Figs 1 and 2. **c**, Identities and fitness

effects of significantly beneficial or deleterious mutations (chromosome number in parenthesis). Asterisks indicate fitness effects measured from reconstructions (mean of six replicate fitness assays, error bars \pm s.e.m.); other fitnesses are from sequencing-based assay (error bars \pm s.e. of regression coefficient; Methods). Italicized mutations are synonymous.

action of balancing selection at the *ERG3* locus. Our earlier work shows that this stable polymorphism can also occur in asexual populations, but much less commonly³⁰, possibly owing to clonal interference limiting the initial spread of ergosterol mutants. Further work is required to fully characterize how interactions between sex and balancing selection affect the evolutionary dynamics and long-term stability of this phenotypic diversification.

Together, our results show that sex increases the rate of adaptation both by combining beneficial mutations into the same background and by separating deleterious mutations from advantageous backgrounds that would otherwise drive them to fixation. In other words, sex makes natural selection more efficient at sorting beneficial from deleterious mutations. This alters the rate and molecular signatures of adaptation. These benefits persist even when balancing selection maintains phenotypic polymorphism within the population. Future studies are needed to fully understand the consequences of this interplay between sex and balancing selection, and to investigate how epistasis interacts with recombination to alter the dynamics of sequence evolution. By combining precise control of the sexual cycle with whole-population time-course sequencing, this experimental system offers the potential to understand how these factors affect the rate, molecular outcomes, and repeatability of adaptation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 May 2015; accepted 20 January 2016.

Published online 24 February 2016.

1. Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Univ. California Press, 1982).
2. Otto, S. P. & Lenormand, T. Resolving the paradox of sex and recombination. *Nature Rev. Genet.* **3**, 252–261 (2002).
3. Kondrashov, A. S. Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**, 372–387 (1993).
4. Weismann, A. in *Essays upon Heredity and Kindred Biological Problems* (eds Poulton, E. B., Schonland, S. & Shipley, A. E.) 251–332 (Clarendon, 1889).
5. Fisher, R. A. *The Genetical Theory of Natural Selection* Ch. 6 (Oxford Univ. Press, 1930).
6. Muller, H. Some genetic aspects of sex. *Am. Nat.* **66**, 118–138 (1932).
7. Peck, J. R. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* **137**, 597–606 (1994).
8. Johnson, T. & Barton, N. H. The effect of deleterious alleles on adaptation in asexual populations. *Genetics* **162**, 395–411 (2002).
9. Gray, J. C. & Goddard, M. R. Sex enhances adaptation by unlinking beneficial from detrimental mutations in experimental yeast populations. *BMC Evol. Biol.* **12**, 43 (2012).
10. Becks, L. & Agrawal, A. F. The evolution of sex is favoured during adaptation to new environments. *PLoS Biol.* **10**, e1001317 (2012).
11. Zeyl, C. & Bell, G. The advantage of sex in evolving yeast populations. *Nature* **388**, 465–468 (1997).
12. Goddard, M. R., Godfray, H. C. J. & Burt, A. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434**, 636–640 (2005).
13. Colegrave, N. Sex releases the speed limit on evolution. *Nature* **420**, 664–666 (2002).

14. Poon, A. & Chao, L. Drift increases the advantage of sex in RNA bacteriophage Φ6. *Genetics* **166**, 19–24 (2004).
15. Becks, L. & Agrawal, A. F. Higher rates of sex evolve in spatially heterogeneous environments. *Nature* **468**, 89–92 (2010).
16. Rice, W. R. & Chippindale, A. K. Sexual recombination and the power of natural selection. *Science* **294**, 555–559 (2001).
17. Cooper, T. F. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol.* **5**, e225 (2007).
18. Weissman, D. B. & Barton, N. H. Limits to the rate of adaptive substitution in sexual populations. *PLoS Genet.* **8**, e1002740 (2012).
19. Crow, J. F. & Kimura, M. Evolution in sexual and asexual populations. *Am. Nat.* **99**, 439–450 (1965).
20. Maynard Smith, J. What use is sex? *J. Theor. Biol.* **30**, 319–335 (1971).
21. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
22. Kao, K. C. & Sherlock, G. Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nature Genet.* **40**, 1499–1504 (2008).
23. Miralles, R., Gerrish, P. J., Moya, A. & Elena, S. F. Clonal interference and the evolution of RNA viruses. *Science* **285**, 1745–1747 (1999).
24. Sella, G., Petrov, D. A., Przeworski, M. & Andolfatto, P. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* **5**, e1000495 (2009).
25. Good, B. H. & Desai, M. M. Deleterious passengers in adapting populations. *Genetics* **198**, 1183–1208 (2014).
26. Schiffels, S., Szöllősi, G. J., Mustonen, V. & Lässig, M. Emergent neutrality in adaptive asexual evolution. *Genetics* **189**, 1361–1375 (2011).
27. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
28. Hartfield, M. & Otto, S. P. Recombination and hitchhiking of deleterious alleles. *Evolution* **65**, 2421–2434 (2011).
29. Birkby, C. W. & Walsh, J. B. Effects of linkage on rates of molecular evolution. *Proc. Natl. Acad. Sci. USA* **85**, 6414–6418 (1988).
30. Frenkel, E. M. et al. Crowded growth leads to the spontaneous evolution of semi-stable coexistence in laboratory yeast populations. *Proc. Natl. Acad. Sci. USA* **112**, 11306–11311 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J.-Y. Leu and S. Akle-Serrano for help with strain construction and experimental evolution; S. Kryazhimskiy, E. Jerison, and J. Piper for help with sequencing library preparation; G. Lang, A. Murray, B. Good, D. van Dyken, K. Kosheleva, I. Cvijović, and other members of the Desai laboratory for discussions and comments on the manuscript; and P. Rogers and C. Daly for technical support. D.P.R. acknowledges support from an NSF graduate research fellowship. M.M.D. acknowledges support from the James S. McDonnell Foundation, the Alfred P. Sloan Foundation, the Harvard Milton Fund, the Simons Foundation (grant 376196), grant PHY 1313638 from the National Science Foundation, and grant GM104239 from the National Institutes of Health. Computational work was performed on the Odyssey cluster supported by the Research Computing Group at Harvard University.

Author Contributions M.J.M., D.P.R., and M.M.D. designed the project; M.J.M. conducted the experiments and generated the sequencing data; D.P.R. designed and conducted the bioinformatics analysis; M.J.M., D.P.R., and M.M.D. analysed the data and wrote the paper.

Author Information Genome sequence data have been deposited in GenBank under BioProject identifier PRJNA308843. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M.D. (mmdesai@fas.harvard.edu).

METHODS

No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Genotype and strain construction. The strains used in this study were derived from the base strains JYL1129 and JYL1130, haploid W303 yeast strains with genotypes *MATa*, *STE5pr-URA3 ade2-1 his3Δ::3xHA leu2Δ::3xHA trp1-1 can1::STE2pr-HIS3 STE3pr-LEU2* and *MATα STE5pr-URA3 ade2-1 his3Δ::3xHA leu2Δ::3xHA trp1-1 can1::STE2pr-HIS3 STE3pr-LEU2* respectively (provided by J.-Y. Leu). Note these strains contain nutrient markers driven by promoters that are specific to haploid cells (*STE5pr-URA3*) and either mating type **a** (*STE2pr-HIS3*) or mating type **α** (*STE3pr-LEU2*)³¹. We identified a likely non-functional open reading frame (*YCR043C*) as an ideal target for insertion of mating-type-specific drug resistance markers close to the *MAT* locus. We amplified flanking regions from genomic DNA obtained from the *YCR043C* deletion mutant of the *S. cerevisiae* whole-genome deletion collection³² using primers KANampFw and KANampRv (Supplementary Data 2) and integrated this product at the *YCR043C* locus of JYL1129 to generate strain MJM64. We then amplified the *HPHB* gene from plasmid pJHK137 (provided by J. Koschwanez) using primers HYGampFw and HYGampRv (Supplementary Data 2) and integrated at the *YCR043C* locus of JYL1130 to generate strain MJM36.

Evolution experiment. We founded 12 mating type **a** lines using strain MJM64 and 12 mating type **α** lines using strain MJM36. Each of our 6 sexual populations consists of one specific pair of these *MATa* and *MATα* lines. The other 6 *MATa* and 6 *MATα* lines were designated as asexual controls (a total of 12 asexual controls). Between sexual cycles, we propagated these lines at 30 °C in unshaken round bottom 96-well plates containing 128 µl of yeast extract peptone dextrose (YPD) with daily 1:2¹⁰ dilutions using a Biomek FX liquid handling robot (Beckman Coulter). Pairs of *MATa* and *MATα* lines representing a single sexual population were propagated independently in this mitotic phase. As previously described³³, this protocol results in approximately ten generations per day and an effective population size of $N_e \approx 10^5$. Aliquots from generation 30 of each 90-generation cycle were mixed with glycerol to 25% and kept at -80 °C for long-term storage.

After each 90 generations of asexual propagation, we initiated sexual cycles in the sexual populations. In each sexual cycle, we mixed and mated each pair of *MATa* and *MATα* lines, sporulated the resulting six diploid populations, isolated **a** and **α** subpopulations, and used these to initiate another 90 generations of mitotic growth (Extended Data Fig. 1). To mate our lines we mixed **a** and **α** haploids, spotted onto YPD plates, and then incubated at 30 °C. After 5 h, cells were scraped from the plate, resuspended in PBS buffer solution and then plated on YPD agar containing hygromycin (300 µg ml⁻¹) and G418 (200 µg ml⁻¹) to select for diploids. For sporulation, 10 µl of saturated diploid culture was inoculated into 1 ml of yeast peptone acetate liquid media for incubation on a roller drum at 21 °C. After 12–15 h, cells were pelleted, resuspended in 1 ml of 1 M KOAc and then incubated at room temperature with agitation in a roller drum. After 3 days, the presence of spores was confirmed by microscope. We then pelleted and resuspended cells in Zymolase solution (Zymo Research, 0.4 U µl⁻¹) to digest spore walls and eliminate the majority of unmated diploids. To ensure that only mated and sporulated individuals survived this treatment, the zymolase lysate was divided, with one half plated onto defined amino-acid dropout media CSM (-uracil, -leucine) to select for **α** haploids, and the other half plated onto CSM (-uracil, -histidine) to select for **a** haploids. After 24 h of growth at 30 °C, the lawn of cells was washed from plates and diluted into liquid CSM (-uracil, -leucine) or CSM (-uracil, -histidine) and propagated for 24 h. We used a dilution series to estimate the population size of this lawn, to confirm that this procedure did not lead to a population size bottleneck compared with the effective population size. Cultures were checked for diploids by plating a sample on YPD containing G418 and hygromycin to quantify the number of unsporulated diploids that survive haploid selection. We found that diploid leakage was never more than 0.1% (see Extended Data Table 1 for details). These cultures were diluted into YPD and propagated for 90 generations before the sexual cycle was repeated. Asexual control populations were maintained in the same conditions as sexuals wherever possible, with the exception of sporulation, during which time these populations were kept at 17 °C without dilution or agitation.

In principle, sexual and asexual populations could adapt differentially to the conditions specific to the sexual and asexual treatments. To test whether this effect could drive any differences between sexual and asexual lines, we measured the relative fitness of all evolved lines compared with the ancestor in both the sporulation and the 17 °C treatment conditions. Specifically, we acclimated six replicates of each evolved strain to YPD for 24 h and then mixed each with a fluorescently marked ancestral strain in equal proportions. We subjected three of these replicate populations of each evolved strain to the 17 °C treatment (plates were sealed and incubated at 17 °C for 4 days) and the other three to the

sporulation treatment (incubation for 1 day in yeast peptone acetate liquid media, followed by 3 days in 1 M KOAc at room temperature). We used flow cytometry (Fortessa, BD Biosciences) to measure the ratio of the two competing types immediately after mixing and again immediately after the 4-day treatment, counting approximately 20,000 cells for each measurement. We found that both sexual and asexual evolved lines performed better than the ancestor in 17 °C treatment and worse in the sporulation treatment (Extended Data Fig. 2). However, the effects of the sporulation and 17 °C treatments did not vary systematically between evolved sexual and asexual populations (two-sided *t*-test, $P = 0.5$ and $P = 0.8$ respectively), and averaged over a 90-generation cycle any differences were small compared with the gains in fitness attained during adaptation to YPD. Thus there is no evidence that adaptation to sporulation or 17 °C played any role in our results.

We also tested whether conditions specific to the asexual treatment (4 days at 17 °C without dilution) or the sexual treatment (4 days of sporulation without dilution) caused variation in the number of mutations that occur in sexual and asexual lines. We assayed mutation rate by counting the number of spontaneous 5-fluoroorotic acid (5-FOA) resistant mutants that arose in independent cultures of the ancestral W303 strain. Specifically, we propagated 54 populations in a microwell plate containing 128 µl YPD. After one dilution cycle, we plated 18 of these cultures on agar plates containing SC-uracil supplemented with 1 mg ml⁻¹ 5-FOA (Sigma/Aldrich), and we counted the number of 5-FOA-resistant mutants in each culture. Of the remaining 36 cultures, we incubated 18 for 4 days at 17 °C in a microplate, and put 18 through our sporulation cycle (1 day in yeast peptone acetate and 3 days in 1 ml of KOAc). We then plated both sets of cultures on selective media and counted the total number of mutants in each (Extended Data Table 2). We then calculated the number of mutations per culture (m) using the Ma-Sandri-Sarkar maximum likelihood method³⁴. We found no difference in the numbers of mutations across all three data sets, suggesting that most mutations occurred primarily during growth in YPD, and not during incubation at 17 °C or during sporulation culture conditions.

We note that each sexual population consists of a mating type **a** and a mating type **α** subpopulation, while each asexual population consists of a single type **a** or type **α** line. Although sexual populations were bottlenecked to the same total size as the asexuals during each sexual cycle, this difference meant there was a potential difference in effective population size between treatments. To test whether this difference could explain the more rapid adaptation in sexual populations, we evolved an alternative set of 6 asexual control populations for 990 generations. Each of these alternative asexual controls consisted of one specific pair of *MATa* lines (that is, two *MATa* subpopulations per asexual population). We propagated these subpopulations separately between sexual cycles. Every 90 generations, we mixed the two subpopulations (exactly analogous to the sexual lines but without recombination) and then divided them for another 90 generations of separate propagation. Simultaneously, we evolved 12 additional asexual control lines propagated in the same manner but without mixing every 90 generations. After 990 generations of evolution, we measured the fitness of all evolved populations. We find these mixed and unmixed asexual controls adapt at the same rate (Extended Data Fig. 3, two-sided *t*-test, $P = 0.8$). Thus this difference in treatments is not responsible for the faster adaptation in sexual populations.

Fitness assays. Fitness assays were performed as described previously³³. Briefly, fitness was measured by competing test clones or populations against an ancestral reference strain containing an mCitrine fluorescent marker inserted at the *HIS3* locus³⁵. Because this reference strain would mate with *MATα* lines, all population fitness assays were performed on *MATa* subpopulations. After strains had acclimated to YPD media for 24 h, competing strains were mixed in equal proportions and propagated by diluting 1:2¹⁰ every 24 h. We used flow cytometry (Fortessa, BD Biosciences) to measure the ratio of the two competing types after 1 and 3 days (approximately 10 generations and 30 generations respectively), counting approximately 20,000 cells for each measurement. We confirmed the appropriateness of each *t*-test conducted using this fitness data with an *F*-test.

Sequencing and variant calling. Glycerol stocks of populations to be sequenced were defrosted and 10 µl inoculated into 3 ml of YPD and incubated without shaking at 30 °C for 16 h (*MATa* and *MATα* subpopulations of each sexual line were sequenced separately). Genomic DNA was prepared from these cultures using a Yeastar Genomic DNA kit (Zymo Research). Library preparations were prepared with a Nextera kit, using a protocol we previously described³⁶. Libraries were sequenced to an approximate depth of 40-fold coverage using an Illumina HiSeq 2500 (Illumina).

We aligned Illumina reads from all samples (after trimming Nextera adaptor sequences) to a SNP/indel-corrected W303 reference genome²¹ using bowtie2 version 2.1.0 (ref. 37). Next, we marked duplicate reads with Picard version 1.44. We generated a list of candidate SNPs and indels by applying GATK's UnifiedGenotyper version 2.3 to all time points in each population at once³⁸. To find low-frequency variants, we set the minimum phred-scaled confidence

threshold for GATK to call a mutation to 4.0. For each candidate mutation, we extracted the allele depth supporting the reference and alternate allele from the resulting VCF file and calculated mutation frequencies for each time point. We excluded potential mutations if there was less than $10 \times$ average coverage across all time points or if GATK called two or more alternate alleles at that site. We required that a mutation be supported by at least ten total reads and that it reach a frequency of 0.1 in two or more time points of the population in which it was called.

To refine our list of candidate mutations, we took advantage of our time-course sequencing and multiple replicate lines. The frequency of a real mutation should be correlated across time points, while errors should be uncorrelated. We thus excluded candidate mutations whose frequency trajectories were uncorrelated (lag-1 autocorrelation less than 0.2). Also, it is unlikely that the same base-pair substitution will arise independently in replicate populations. Thus, for each candidate mutation, we estimated the site-specific error rate by calculating the frequency of the alternate allele outside of the population in which the mutation was called. We then excluded candidates with an estimated error rate above 0.05. We also calculated the probability of detecting at least the observed number of alternate alleles in the focal population, assuming a binomial error model (given the observed coverage and estimated error rate). We excluded candidates where this probability exceeded 10^{-5} . We also detected several mutations that were present in the founding stock and thus in multiple replicate populations. We marked these mutations and excluded them from our counts of *de novo* mutations and from Fig. 2. After performing this procedure in the *MATA* and *MATα* subpopulations of each sexual line separately, we combined called mutations from both subpopulations and averaged the mutation frequencies to generate the whole-population trajectories in Fig. 2; data on each subpopulation separately are available in Supplementary Data 1.

We annotated each called mutation using a SNP/indel-corrected GFF file and determined its effect on amino-acid sequence. We also screened for complex mutations: pairs of mutations that were within 1 kb of one another and followed the same trajectory. We discovered 7 complex mutations, all within 41 bases of one another. We determined the net effect of each complex mutation and considered them to be single mutations in our analysis.

We note that it is not possible to determine the fraction of mutations that we detect with our variant-calling method. For example, sequencing depth fundamentally limits our ability to detect rare mutations. We do not attempt to call mutations that never reach $\sim 10\%$ frequency because our 40-fold coverage gives no resolution below that level; our results thus represent only mutations that reach substantial frequency. We are also limited to the set of mutations that can be identified by GATK, mainly SNPs and small indels (but see below for an analysis of larger-scale mutations from clone sequence data). These limitations apply equally to our sexual and asexual populations.

Correlations between frequency trajectories. Clonal interference is expected to generate correlations between the frequency trajectories of mutations that segregate at the same time. Two mutations in the same genetic background should increase or decrease together, while mutations on different backgrounds will tend to move in opposite directions. For each mutation trajectory, we calculated the change in frequency between each sequenced time point. We then computed the correlation coefficient between changes in the same time interval for every pair of mutations in the same population. We excluded pairs of mutations that did not segregate at the same time (that is, pairs whose frequencies were never between 0.05 and 0.95 in the same time point). Because large positive and large negative correlation coefficients are both evidence of interference effects, we compared the distributions of squared correlation coefficients (R^2) in asexual and sexual populations (Fig. 2i–l).

The dynamics of natural selection will introduce such correlations even among unlinked mutations by constraining the shapes of frequency trajectories. For example, two simultaneous but genetically unlinked selective sweeps will each follow a similar sigmoidal trajectory and thus be strongly correlated with one another. We controlled for this effect by repeating the above calculations with all pairs of mutations segregating in different populations of the same reproductive type. The R^2 values from this procedure comprise two empirical null distributions (sexual and asexual) for mutations that are certain to be independent of one another (Fig. 2j–l). **Detection of large deletions and copy-number variants.** Our primary variant-calling pipeline can only detect substitutions, insertions, and deletions affecting ~ 3 bp or less. To estimate the prevalence of larger-scale mutations in our populations, we implemented an alternative pipeline to detect large deletions and copy-number variants on the basis of coverage depth as a function of genome position. Coverage depth in whole-population samples is difficult to interpret because it convolves individual copy-number with population variation. For example, a fixed duplication and a fourfold amplification present in half the population would generate identical coverage data in a whole-population sample. To avoid this problem, we sequenced eight total clones isolated from the final time points of two sexual and two asexual populations to an average depth per clone of 50–80×.

After aligning reads to the reference as described above, we tabulated coverage depth in 100 base-pair windows as the number of mapped reads whose start positions fell within each window. These windows vary naturally in coverage depth owing to pre-existing duplications, PCR artefacts, and properties of the alignment algorithm. Therefore, to generate a baseline expectation, we calculated coverage in the same windows for all of the generation-0 and generation-90 population samples. Added together, these data yielded 564 reads in the median window. We thus calculated the expected relative coverage in each window by dividing its total coverage in the generation-0 and generation-90 samples by 564. For each clone, we then multiplied this expected relative coverage by the median coverage per window in that clone to get the expected coverage in each window.

We next looked for windows in which the observed coverage depth deviated from its expectation. This is complicated by the fact that random noise is introduced by the sequencing and alignment process. Because the coverage depth is generated by a counting process, the noise variance scales with the expected coverage. We therefore applied a variance-stabilizing Anscombe transform³⁹ to standardize the noise across windows with different expectations. First, we modelled the variance as $v(m) \propto m + m^2/r$, where m is the expected coverage in a window, v is the mean squared deviation from that expectation, and r is a parameter fit to the data by a linear regression of v/m by m (we find a best-fit value $r \approx 440$). This variance function, which is characteristic of negative-binomial counting noise, leads to an Anscombe transformation $A(k) = \text{arcsinh} \left[\sqrt{\frac{k+c}{r-2c}} \right]$, where k is the observed coverage and $c = 3/8$ following the recommendation of ref. 39 for negative-binomial data. The transformed data are approximately normally distributed with mean $A(m)$ and constant variance.

Deletions and amplifications larger than our 100-bp window size should generate spatially correlated signals in our data, while the variance-stabilized noise will be largely uncorrelated between adjacent windows. To take advantage of this, we performed a ‘wavelet denoising’ procedure, a standard signal-processing method for separating spatially correlated signals from white noise⁴⁰, which has been used previously⁴¹ in similar analyses of biological sequence data. Specifically, we applied a discrete wavelet transform with the Haar basis, using the Python package PyWavelets, to our variance-stabilized and mean-centred data. We then performed noise reduction by replacing each wavelet coefficient a_i with a thresholded coefficient a_i^* , according to the formula $a_i^* = \text{sign}(a_i) \max[0, |a_i| - t]$, where the threshold value t was set to three standard deviations of the variance-stabilized data.

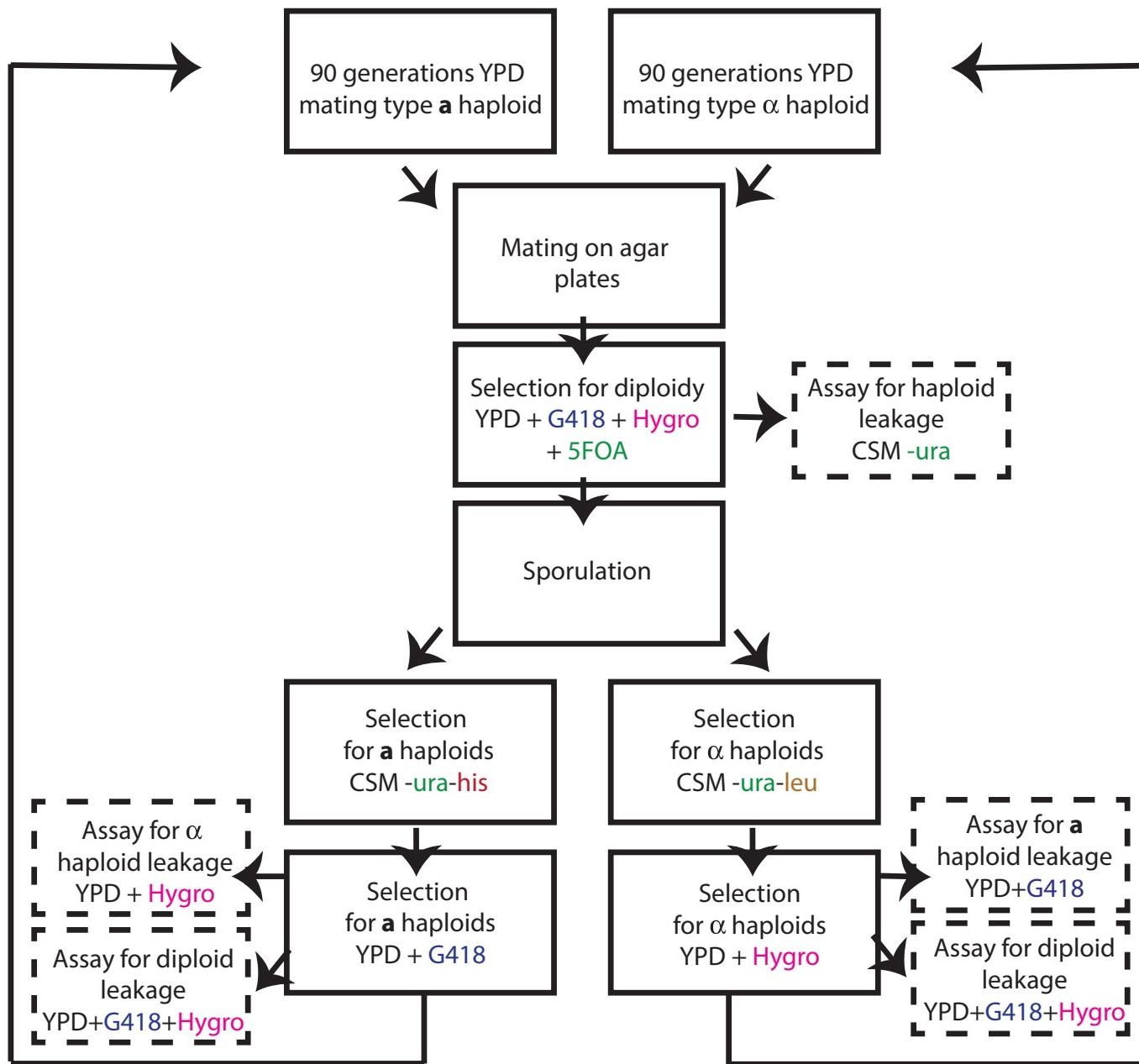
After noise reduction, we inverted the wavelet and Anscombe transforms to get a smoothed estimate of the ratio of observed to expected coverage as a function of position (Extended Data Fig. 4). By visual inspection, we identified ten regions exhibiting strong signals of amplification or deletion in at least one clone (Extended Data Table 4). Of these, two regions (an rDNA-rich segment of chromosome XII and the segment of chromosome VIII containing CUP1-1 and CUP1-2) seemed to have undergone amplification in multiple independent populations. Both of these regions are known to exhibit copy-number variation across *S. cerevisiae* strains^{42,43}. Of the remaining regions, five contained Ty elements.

Genetic dissection and reconstructions. To probe their fitness effects, we reconstructed mutations from evolved strains in the mating type **a** ancestral genetic background, *MATA*, *ura3Δ::NATMX*, *ade2-1*, *his3Δ::3xHA*, *leu2Δ::3xHA*, *trp1-1*, *CAN1*. First, DNA fragments containing *URA3* and *HPHB* were amplified from plasmid pJHK137 using primers containing 40 nucleotides of homology to sequence on each side of the target nucleotide (see Supplementary Data 2 for primer sequences). The mating type **a** ancestor was transformed with the resulting PCR product, resulting in hygromycin-resistant *URA*⁺ strains. These mutants were in turn transformed with an 80-bp double-stranded oligonucleotide centred on the mutant allele (see Supplementary Data 2). We plated on 5-FOA to select for the replacement of the *URA3* genes with the mutant allele, and confirmed replacement by replica plating on YPD + hygromycin. Correct genotypes were confirmed by Sanger sequencing.

We found one example of a mutation in *MET2* that had a strong deleterious effect when introduced into the ancestral genetic background, despite fixing in a sexual population. We also found that this mutation had no significant effect in the sequencing fitness assay. To investigate whether epistasis could be responsible for these observations, we sought to measure the effect of this *met2* mutation in the evolved background from the sexual population in which it fixed. To use our *URA3-HPHB* strategy, we first replaced the *STE5pr::URA3* locus in the evolved clone with a *NATMX* marker, resulting in *ura3Δ::NATMX* (primers in Supplementary Data 2). We confirmed that this manipulation did not affect fitness. We then used this strain as the basis to reintroduce the wild-type *MET2* allele. The resulting difference in fitness between the evolved sexual clone and reconstructed wild type was used to calculate the fitness effect for the *met2* allele shown in Fig. 3c.

Sequencing-based fitness assay. To measure the fitness effects of mutations in evolved populations, we sampled a single evolved clone from generation 990 of each of the sequenced asexual populations and from generation 630 of each of the sequenced sexual populations. We backcrossed each of these clones with its corresponding ancestor. This resulted in diploids heterozygous for all mutant sites that were present in each original clone. We bulk sporulated each of these diploids to generate a large number of recombinant haploids with different combinations of wild-type and mutant alleles. Each of these populations of haploids was then propagated in YPD liquid medium in the same conditions used during mitotic propagation in the evolution experiment. We sampled each population after 10, 30, 50, and 70 generations, prepared genomic DNA, and sequenced to measure the frequencies of each mutation over time. We estimated the fitness effect of each mutation (Fig. 3 and Supplementary Data 1) from the coverage depth supporting the mutant and ancestral alleles as a function of time (binomial regression with a logistic link function, coefficients and standard errors calculated using the *glm* function in R).

31. Tong, A. H. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
32. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
33. Lang, G. I., Botstein, D. & Desai, M. M. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* **188**, 647–661 (2011).
34. Hall, B. M., Ma, C.-X., Liang, P. & Singh, K. K. Fluctuation Analysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis. *Bioinformatics* **25**, 1564–1565 (2009).
35. Frenkel, E. M., Good, B. H. & Desai, M. M. The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. *Genetics* **196**, 1217–1226 (2014).
36. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014).
37. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
38. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).
39. Anscombe, F. J. The transformation of poisson, binomial and negative-binomial data. *Biometrika* **35**, 246–254 (1948).
40. Donoho, D. L. & Johnstone, I. M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **90**, 1200–1224 (1995).
41. Shim, H. & Stephens, M. Wavelet-based genetic association analysis of functional phenotypes arising from high-throughput sequencing assays. *Ann. Appl. Stat.* **9**, 665–686 (2015).
42. Dunham, M. J. *et al.* Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **99**, 16144–16149 (2002).
43. Chang, S.-L., Lai, H.-Y., Tung, S.-Y. & Leu, J.-Y. Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet.* **9**, e1003232 (2013).

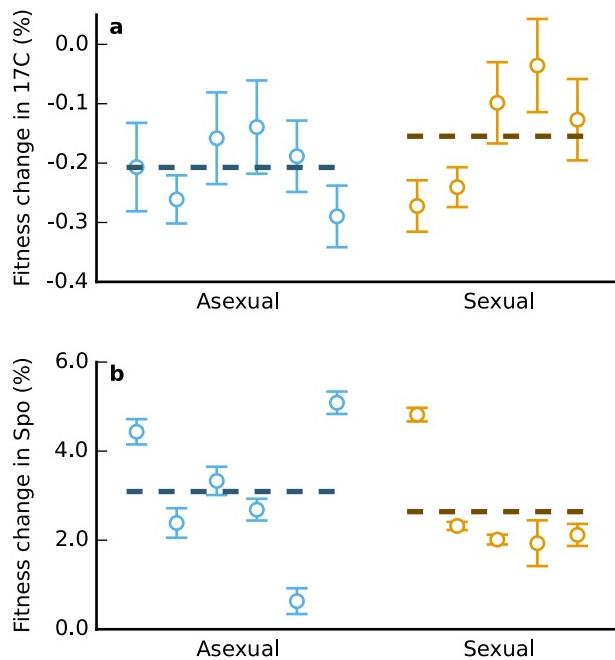


MAT α -YCR043C::KANMX STEP5pr-URA3, CAN::STE2pr-HIS3, STE3pr-LEU2

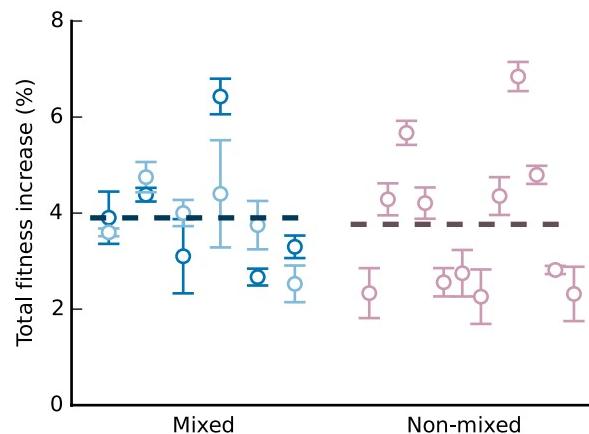
MAT α -YCR043C::HPHB STEP5pr-URA3, CAN::STE2pr-HIS3, STE3pr-LEU2

Extended Data Figure 1 | Genetic system and experimental protocol for evolution of sexual populations. Genotypes of the two haploid mating types are indicated at bottom, with selectable markers that are expressed in each strain indicated in colour. Steps in our experimental protocols involving these markers are indicated in the corresponding colour. *STE5pr* is a haploid-specific promoter and *STE2pr* and *STE3pr* are α - and

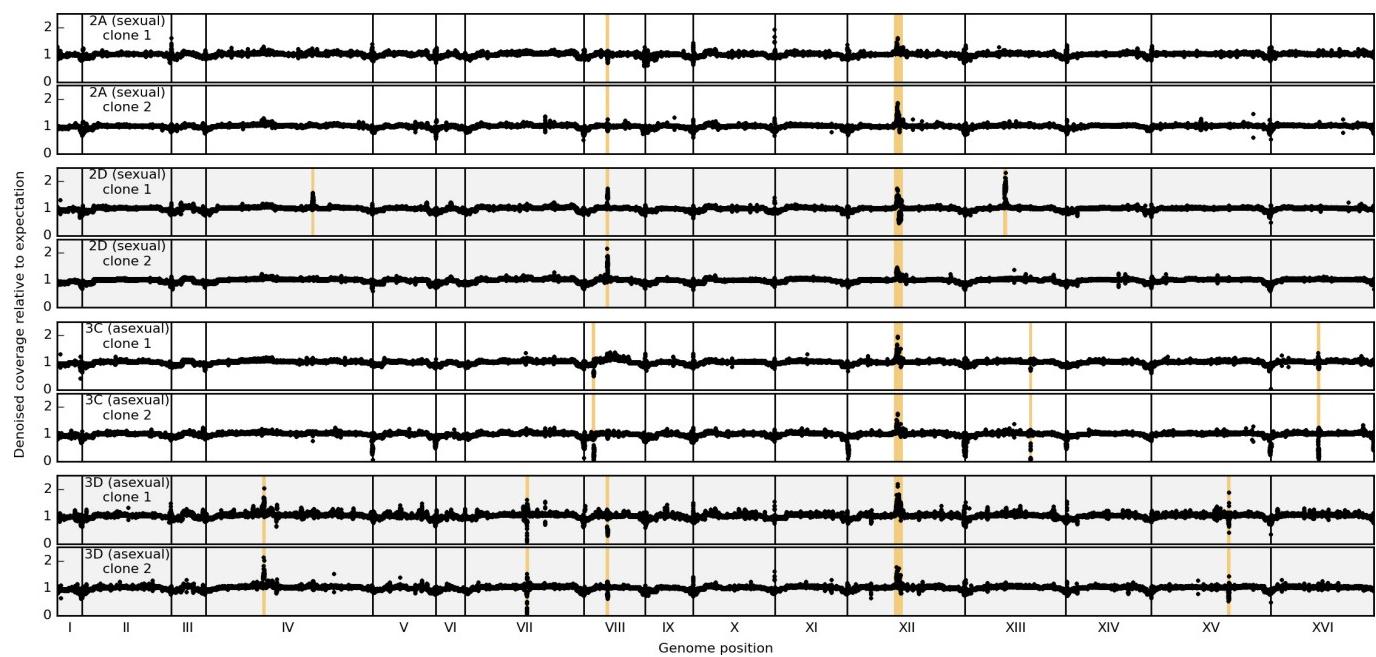
α -specific promoters respectively, so haploid α cells express *URA3* and *HIS3*, while haploid α cells express *URA3* and *LEU2*. The drug resistance markers *KANMX* and *HPHB*, tightly linked to the α and α mating loci respectively, are constitutively expressed. *URA3* is counterselectable; it is not expressed in diploids, rendering them resistant to 5-FOA.



Extended Data Figure 2 | Adaptation to 17 °C and sporulation conditions. **a, b**, Relative fitness of evolved asexual (blue) and sexual (orange) populations over four days in 17 °C (**a**) and sporulation conditions (**b**). Fitness changes are reported averaged over a complete experimental cycle (90 generations; mean of three replicate fitness assays, error bars \pm s.e.m.). Mean fitness differences between asexual and sexual evolved strains are not significant in either the 17 °C (two-sided *t*-test, $P = 0.5$) or sporulation (two-sided *t*-test, $P = 0.8$) treatment.



Extended Data Figure 3 | Adaptation in mixed and non-mixed asexual populations. Fitness increases after 990 generations of evolution in mixed (blue) and non-mixed (pink) alternative asexual control populations (mean of four replicate fitness measurements, error bars \pm s.e.m.). Each non-mixed line was maintained independently. Subpopulations from mixed populations were mixed in pairs every 90 generations; each pair is indicated by a corresponding light and dark circle.



Extended Data Figure 4 | Read-depth variation analysis of sequenced clones. Denoised, normalized coverage in 100-bp windows along the genome (Methods). Each panel represents a clone isolated from one of four independent populations. Pairs of clones from the same population

are adjacent and indicated by the population label on the left. Regions containing putative amplifications and deletions (Extended Data Table 4) are highlighted in orange.

Extended Data Table 1 | Leakage of diploids through the sexual cycle

Population	Generation										
	0	90	180	270	360	450	540	630	720	810	990
2A	<0.1	<0.1	0.8	<0.1	0.7	<0.1	0.7	0.1	0.6	0.4	0.9
2B	<0.1	<0.1	0.9	0.2	0.2	<0.1	2.4	0.1	0.9	0.9	<0.1
2C	<0.1	<0.1	1.0	<0.1	<0.1	<0.1	2.0	0.1	0.2	0.2	<0.1
2D	<0.1	<0.1	4.0	0.4	<0.1	<0.1	1.7	0.3	0.1	0.5	<0.1
2E	<0.1	<0.1	2.0	<0.1	0.6	<0.1	2.1	0.5	0.4	0.5	<0.1
2F	<0.1	<0.1	3.0	0.1	0.8	<0.1	0.8	0.1	0.1	0.1	<0.1
5A	<0.1	6.0	<0.1	0.1	<0.1	<0.1	<0.1	<0.1	0.2	<0.1	0.5
5B	<0.1	3.0	<0.1	0.6	<0.1	<0.1	1.1	<0.1	<0.1	<0.1	1.1
5C	<0.1	4.5	<0.1	0.5	<0.1	<0.1	0.6	<0.1	<0.1	<0.1	0.8
5D	<0.1	<0.1	<0.1	1.7	<0.1	<0.1	0.4	<0.1	0.4	<0.1	0.3
5E	<0.1	6.0	<0.1	0.4	<0.1	<0.1	0.2	<0.1	<0.1	0.1	0.4
5F	<0.1	4.5	<0.1	0.1	<0.1	<0.1	0.1	<0.1	<0.1	<0.1	0.1

Fraction ($\times 10^4$) of diploid leakage observed in each sexual population after sporulation, immediately before the 90-generation asexual cycle.

Extended Data Table 2 | Mutation frequency in YPD, sporulation and 17 °C treatments

Culture	YPD	YPD+Spo	YPD+17°C
1	1	9	0
2	3	0	18
3	11	1	10
4	7	5	12
5	0	0	34
6	9	15	15
7	9	10	2
8	2	2	1
9	6	2	9
10	3	8	23
11	10	1	0
12	1	4	6
13	2	11	6
14	4	12	1
15	1	9	2
16	9	12	12
17	14	21	1
18	3	2	9
<i>m</i>	3.2	3.2	3.3

Colony counts of 5-FOA-resistant mutants are listed for each treatment. The number of mutations per culture (*m*) was calculated from the colony counts shown using the Ma–Sandri–Sarkar maximum likelihood method.

Extended Data Table 3 | Classification of observed mutations

	All	Nonsyn	Syn	Intergenic
Asexual	All	183	111	27
	Fixed	143 (78%)	88 (79%)	20 (74%)
Sexual	All	167	98	22
	Fixed	27 (16%)	22 (22%)	0 (0%)
				5 (11%)

The total number of mutations observed and fixed in the four sequenced asexual populations and four sequenced sexual populations. We classified mutations as fixed if they attained a frequency greater than 0.8 at the final sequenced time point. The percentage of mutations that were fixed in a given class is shown in parentheses next to the number of fixed mutations.

Extended Data Table 4 | Larger-scale mutations

Chromosome	Start (kb)	End (kb)	Clones	Annotation
ChrIV	525	545	3D-1, 3D-2	ENA5, ENA2, ENA1
ChrIV	975	990	2D-1	Ty
ChrVII	560	575	3D-1, 3D-2	Ty
ChrVIII	80	95	3C-1, 3C-2	Ty
ChrVIII	205	220	2A-1, 2A-2, 2D-1, 2D-2, 3D-1, 3D-2	CUP-1, CUP-2
ChrXII	435	500	2A-1, 2A-2, 2D-1, 2D-2, 3C-1, 3C-2, 3D-1, 3D-2	rDNA
ChrXIII	355	375	2D-1	NUP116, CSM3, ERB1
ChrXIII	595	605	3C-1, 3C-2	ALD3, ALD2
ChrXV	700	715	3D-1, 3D-2	Ty
ChrXVI	430	445	3C-1, 3C-2	Ty

Summary of mutations identified by read-depth variation analysis of sequenced clones. We report the approximate start and end position of each mutation and the specific functional elements affected by each event.

MARCKS-like protein is an initiating molecule in axolotl appendage regeneration

Takuji Sugiura^{1,2†}, Heng Wang³, Rico Barsacchi², Andras Simon³ & Elly M. Tanaka^{1,2†}

Identifying key molecules that launch regeneration has been a long-sought goal. Multiple regenerative animals show an initial wound-associated proliferative response that transits into sustained proliferation if a considerable portion of the body part has been removed^{1–3}. In the axolotl, appendage amputation initiates a round of wound-associated cell cycle induction followed by continued proliferation that is dependent on nerve-derived signals^{4,5}. A wound-associated molecule that triggers the initial proliferative response to launch regeneration has remained obscure. Here, using an expression cloning strategy followed by *in vivo* gain- and loss-of-function assays, we identified axolotl MARCKS-like protein (MLP) as an extracellularly released factor that induces the initial cell cycle response during axolotl appendage regeneration. The identification of a regeneration-initiating molecule opens the possibility of understanding how to elicit regeneration in other animals.

To identify a regeneration-initiating molecule in the salamander *Ambystoma mexicanum* (axolotl), we aimed to functionally screen^{6,7} axolotl complementary DNAs using an *in vitro* salamander myotube cell cycle re-entry assay (*Notophthalmus viridescens*; newt)⁸ with the aim of performing *in vivo* analysis in the axolotl that is convenient for molecular analysis^{9–11}. To establish if axolotl blastema tissue expresses a myotube cell-cycle-entry inducing factor, we injected *Xenopus* oocytes with messenger RNAs from tail blastema, limb blastema or mature limb and assayed the extracellular media on myotubes (Fig. 1a). Tail or limb blastema mRNAs scored positively, comparable to serum, whereas the mature tissue mRNAs showed little inducing activity. We next screened an arrayed 6-day tail blastema cDNA eukaryotic expression vector library for the activity¹². Transfection of DNA representing the entire library as a single pool into HEK293 cells (Fig. 1b, sample WL) yielded cell media that stimulated myotube cell cycle entry (Fig. 1b). This library was fractionated into 12 superpools, which yielded four positive superpools (superpool numbers 6, 9, 10 and 12; Fig. 1b and Extended Data Fig. 1a–f). Sib-selection of superpool 9 through three subfractionation steps resulted in identification of a single clone responsible for the activity (Extended Data Fig. 2a–c).

The positive clone encoded a 224-amino-acid protein containing three conserved domains (myristoylated N terminus, MARCKS homology domain and effector domain) similar to MLP (Extended Data Fig. 3a) showing 74.1%, 68.0% and 80.0% amino acid sequence identity to human MLP (also known as MARCKSL1), including the glycine G2 in the myristoylated domain and two serines in the effector domain (S94 and S105) important for plasma membrane binding (for review see ref. 13). The C-terminal region showed low (14.4%) sequence conservation. Phylogenetic analysis showed that the axolotl sequence clustered with other vertebrate MLPs (Extended Data Fig. 3b).

Previous work in other species has indicated that MLP is an intracellular substrate for protein kinase C (PKC) associated with the plasma

membrane, phagocytic vesicles and actin, while phosphorylation by PKC induced dissociation to the cytoplasm (for review see ref. 14). We asked whether axolotl MLP (AxMLP) acted on myotubes as a secreted factor or whether it induced expression of a secreted factor in the HEK293 cells. To determine if AxMLP was extracellularly released, we transfected an expression plasmid encoding an AxMLP-C-terminal fusion with enhanced GFP (eGFP) or the pEGFP-N1 control construct into HEK293 cells (Extended Data Fig. 3c). Increasing levels of GFP fluorescence intensity were observed in *AxMlp-eGfp*-transfected but not *eGfp-N1*-transfected culture media (Fig. 1c). The percentage of GFP⁺ cells and the total cell number in *AxMlp-eGfp*- and *eGfp-N1*-transfected samples remained equivalent over time (Extended Data Fig. 3d, e). Time-lapse imaging further showed that *AxMlp-eGfp*-transfected cells grew similarly to the control cells (Supplementary Videos 1 and 2), indicating that extracellular AxMLP did not derive from dying cells. When comparing expression of AxMLP to zebrafish, *Xenopus*, mouse, newt and human MLPs in HEK293 cells, we observed that the AxMLP yielded a higher proportion of extracellular protein compared with other species (Fig. 1d). Bioassay of the axolotl versus newt MLP media induced a myotube response corresponding to the amount of protein seen by western blot (Fig. 1e).

To establish necessity, we exposed AxMLP-containing culture media to a polyclonal antibody against AxMLP that inhibited the myotube response, indicating that extracellular AxMLP is required for the activity (Fig. 1f and Extended Data Fig. 3f). To test sufficiency, we purified AxMLP-His, which displayed a characteristic high gel mobility (Extended Data Fig. 3g, h; for review see ref. 13). Exposure of myotubes to purified AxMLP in serum-free conditions yielded a robust myotube response with an approximate half-maximal response at 50.5 ng μ l⁻¹ (Fig. 1g and Extended Data Fig. 2d–f). We conclude that extracellular AxMLP is sufficient to induce myotube cell cycle re-entry.

To determine the *in vivo* function of extracellular AxMLP, we first queried whether purified AxMLP protein injected into uninjured axolotl tail (Fig. 2) and limb (Extended Data Fig. 4) tissue was sufficient to induce cell cycle re-entry. We injected 270 ng of AxMLP followed by injection of 5-bromodeoxyuridine (BrdU) at 3 days post-amputation (dpa) (Fig. 2a and Extended Data Fig. 4e). AxMLP-injected tails contained significantly more BrdU-positive cells (18.9 ± 2.59%) than control tails injected with media depleted of AxMLP (flow-through, 3.20 ± 0.863%; PBS, 3.04 ± 1.00%) (Fig. 2b–d). AxMLP injection caused increased BrdU uptake in all counted cell types in limbs and tails except for myocyte enhancer factor 2C (MEF2C)⁺ muscle nuclei (Fig. 2b–d and Extended Data Fig. 4a–d, f–n). Interestingly, it was recently found that muscle fibres can dedifferentiate during newt limb regeneration, but not in axolotl¹⁵. The responsiveness of axolotl PAX7⁺ satellite cells but not MEF2C⁺ muscle nuclei to AxMLP corresponds with PAX7⁺ satellite cells being the main contributors to muscle regeneration in axolotl¹⁵.

¹DFG Research Center for Regenerative Therapies (CRTD), Technische Universität Dresden, Fetscherstrasse 105, 01307 Dresden, Germany. ²Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany. ³Karolinska Institute, Department of Cell and Molecular Biology, Centre of Developmental Biology for Regenerative Medicine, SE-171 77 Stockholm, Sweden. [†]Present address: DFG Research Center for Regenerative Therapies, Technische Universität Dresden, Fetscherstrasse 105, 01307 Dresden, Germany.

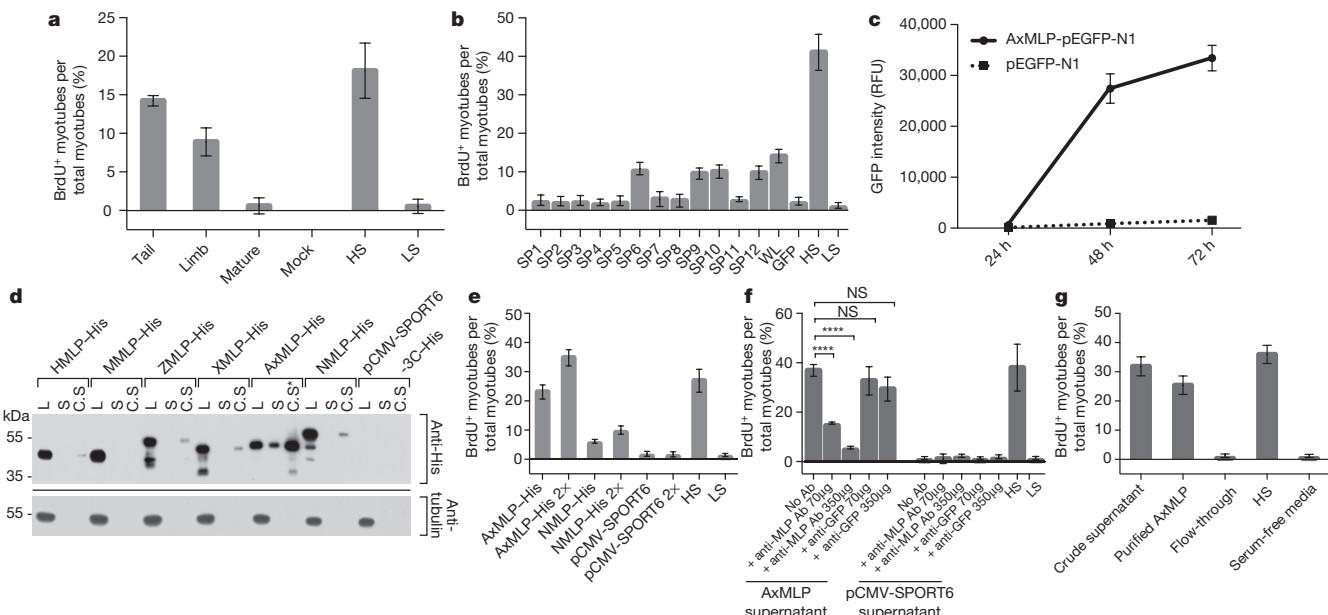


Figure 1 | Extracellular AxMLP identified by expression cloning is necessary and sufficient for cell cycle re-entry *in vitro*.

a, Supernatants from *Xenopus* oocytes injected with total mRNA from tail and limb blastema induced robust BrdU incorporation in cultured newt myotubes (myotube assay) ($n = 6$: 2 biological, 3 technical replicates each; mean \pm standard deviation (s.d.)). b, A screen for the cell-cycle-inducing clone. Culture media from HEK293 cells transfected with 6-day tail blastema cDNA library pools and assayed on myotubes for cell cycle induction (see Extended Data Fig. 1 for scheme) identified four positive superpools ($n = 12$: 4 biological, 3 technical replicates each; mean \pm s.d.). Superpool 9 was sub-selected to a single clone (see Extended Data Fig. 2a–c). c, AxMLP-GFP fusion protein detection in culture media of transfected HEK293 cells by fluorescence luminometry ($n = 3$: biological replicates; mean \pm s.d.). RFU, relative fluorescence units. d, MLP orthologues show

As we had used the newt myogenic cell line for the original screen, we asked whether AxMLP could promote *in vivo* cell cycle entry during muscle dedifferentiation in the newt. AxMLP protein

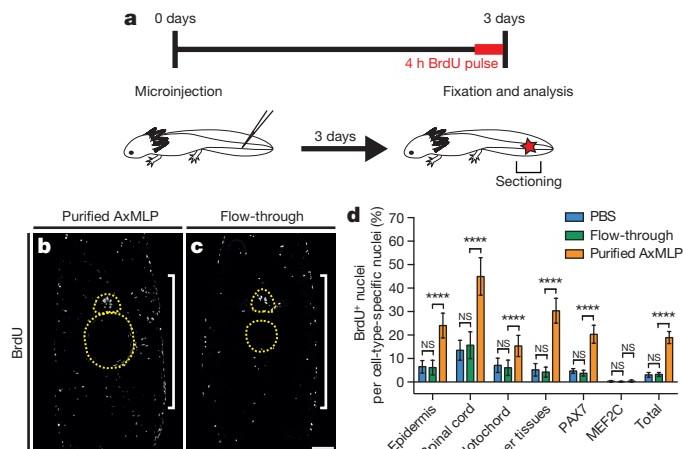


Figure 2 | AxMLP is sufficient to induce cell cycle entry *in vivo*.

a, Schematic illustration of *in vivo* protein injection experiment. b, c, Transverse sections of tails injected with purified AxMLP (b) or flow-through (fraction depleted of AxMLP) (c) immunostained for BrdU. Yellow circles indicate spinal cord (top) and notochord (bottom); white brackets indicate injection site. Scale bar, 200 μ m. d, Quantification of BrdU⁺ cells in injected tails. Quantification of BrdU⁺/PAX7⁺ cells and BrdU⁺/MEF2C⁺ cells shows that AxMLP induces cell cycle entry in PAX7⁺ cells. NS, not significant; ***P < 0.0001 with Student's *t*-test ($n = 15$: 5 biological, 3 technical replicates each; mean \pm s.d.).

differing levels of extracellular protein. AxMLP was readily detectable by western blotting in cell culture supernatant (S). Human (HMLP), zebrafish (ZMLP), *Xenopus* (XMLP) and newt (NMLP) MLP were only detectable in 40-fold concentrated supernatants (C.S.). Asterisk indicates fivefold concentrated supernatant. Loading control: anti-tubulin. e, AxMLP and NMLP supernatants both induce myotube cell cycle with response corresponding to protein levels in supernatant ($n = 6$: 2 biological, 3 technical replicates each; mean \pm s.d.). f, Induction of myotube cell cycle re-entry by AxMLP is specifically blocked by addition of polyclonal anti-AxMLP antibodies (Ab) to culture supernatant ($n = 6$: 2 biological, 3 technical replicates each; mean \pm s.d.). g, Purified AxMLP induces myotube cell cycle re-entry ($n = 6$: 2 biological, 3 technical replicates each; mean \pm s.d.). HS, high serum; L, cell lysate; LS, low serum; SP, superpool; WL, whole library. ***P < 0.0001 with Student's *t*-test. NS, not significant.

was injected either into uninjured newt limbs or after limb amputation during the muscle dedifferentiation phase (Fig. 3). Injection of AxMLP into uninjured newt tissue was not sufficient to induce a cell cycle response (Fig. 3a). Injection during regeneration, however, resulted in an increased 5-ethynyldeoxyuridine (EdU) uptake in PAX7⁺ satellite cells as well as dedifferentiating myofibre-derived cells (Fig. 3b, c)¹⁵. These data indicate that AxMLP can also promote cell cycle entry of at least two cell types in newt, including dedifferentiating muscle. The requirement for an additional injury signal to induce cell cycle entry in newt correlates with a higher propensity of axolotl stem cells to cycle in homeostasis compared with their newt counterparts^{16,17}.

We next asked if AxMLP is important for cell proliferation during axolotl regeneration. Microarray and quantitative reverse transcription polymerase chain reaction (qRT-PCR) detected expression in mature limb and tail tissue followed by upregulation with a peak of expression at 12 to 24 h post-amputation (hpa), returning to basal levels at 2 dpa in the tail and 4 dpa in the limb (Extended Data Fig. 5a, e). These observations are consistent with a role for AxMLP in early events of regeneration. Protein localization using immunofluorescence showed that in uninjured tissue, AxMLP was cytoplasmically localized in epidermis and in tail spinal cord cells, including radial glia and axonal tracts (Extended Data Fig. 5b, f, high-magnification images). At 1 and 6 dpa, expression was maintained in the epidermis and spinal cord (Extended Data Fig. 5c, d, g, h, high-magnification images). However, the protein in the regenerating wound epidermis at 1 dpa was plasma-membrane associated (Extended Data Fig. 5c, g, green arrowheads in high magnification images). Such localization changes have previously been described for MARCKS proteins and are dependent on phosphorylation state¹⁸. In summary, AxMLP protein is

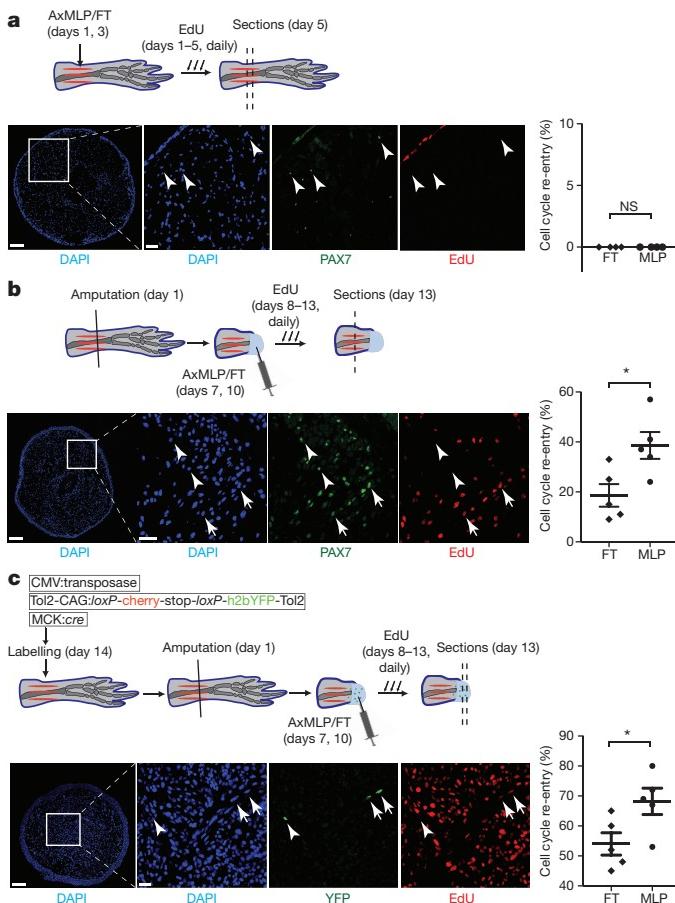


Figure 3 | AxMLP induces cell cycle re-entry of muscle-derived cells in the newt limb. **a–c,** Top, schematic representation of the experimental paradigm testing the effect of AxMLP on PAX7⁺ satellite cell proliferation in uninjured limb (**a**), on PAX7⁺ satellite cell activation in injured limb (**b**) or myofibre dedifferentiation after injury (**c**). **a,** Bottom, transverse section of uninjured limb injected with purified AxMLP shows no induction of EdU incorporation. Graph showing no difference in either purified AxMLP- or flow-through- (FT) injected uninjured limbs (right; $n = 4$; biological replicates). **b,** Bottom, transverse section of the regenerating limb injected with AxMLP shows increased EdU incorporation of PAX7⁺ cells. Graph showing more EdU⁺/PAX7⁺ satellite cells in AxMLP-injected limbs (right; $n = 5$; biological replicates; mean \pm standard error of the mean (s.e.m.)). **c,** Bottom, transverse section from myofibre-labelled, regenerating limb injected with AxMLP. Graph showing more EdU⁺/yellow fluorescent protein (YFP)⁺ myofibre progeny in AxMLP-injected blastemas (right; $n = 5$; biological replicates; mean \pm s.e.m.). NS, not significant; * $P < 0.05$ with Student's *t*-test. Scale bars in lower-magnification images, 200 μm ; in higher-magnification images, 20 μm . Arrowheads indicate marker⁺/EdU[−] cells; arrows indicate marker⁺/EdU⁺ cells. DAPI, 4',6-diamidino-2-phenylindole.

cytoplasmically localized in mature tissue. Upon amputation, mRNA levels rise by at least eightfold. Concomitantly the AxMLP protein in the wound epidermis shows juxtamembrane localization, consistent with its N-terminal myristylation sequence and suggestive of extracellular release. These data suggest that both the level and intracellular localization are critical in the role of AxMLP as a non-autonomous inducer of initial cell cycles. A detailed understanding of the different cytoplasmic pools and their relationship to the extracellular form will require further study.

To test the function of endogenous AxMLP during regeneration, we implemented two different fluorescein isothiocyanate (FITC)-conjugated morpholinos directed against 5' sequences of the *AxMlp* mRNA. We validated the effectiveness of the morpholinos *in vitro* by co-electroporation with plasmid encoding full-length *AxMlp* or

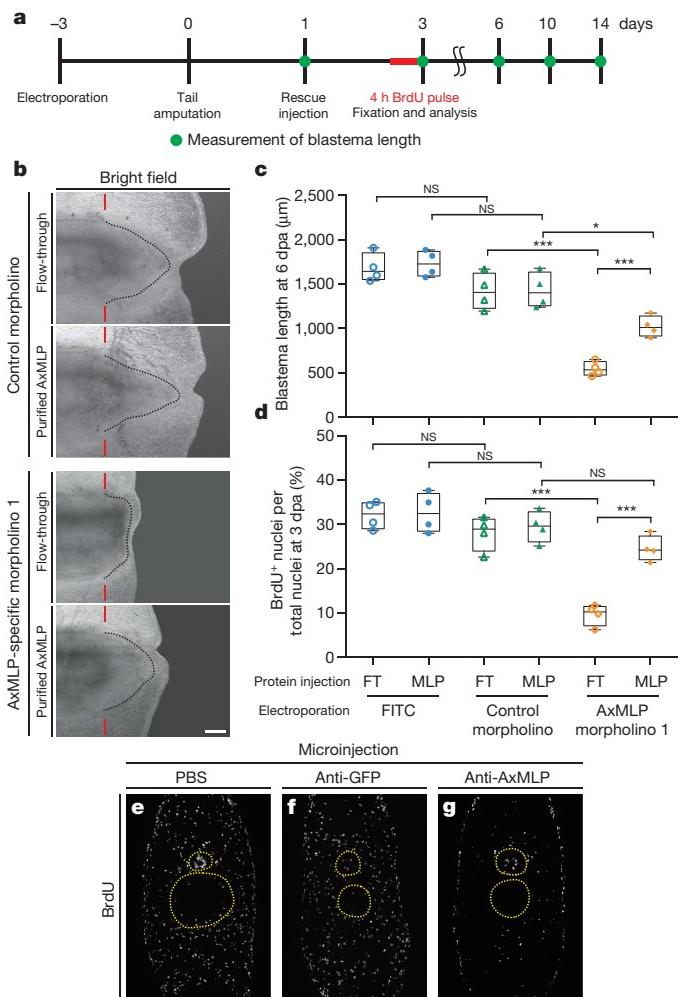


Figure 4 | AxMLP is necessary for cell proliferation during early tail regeneration. **a,** Schematic diagram of the morpholino electroporation experiment. **b,** Bright-field images of the morpholino-electroporated/protein-injected tails at 6 dpa showing inhibition of regeneration by anti-*AxMlp* morpholino and rescue by protein injection. Scale bar, 500 μm . Red bars indicate amputation planes; dashed lines delineate shape of the blastema. **c,** Blastema length at 6 dpa. The sample order on the x-axis is the same as in **d** ($n = 4$; biological replicates; centre values as median; points represent each sample). FT, flow-through. **d,** Quantification of BrdU⁺ cells in tail blastema sections at 3 dpa ($n = 4$; biological replicates; centre values as median; points represent each sample). **e–g,** Injection of anti-AxMLP antibody inhibits proliferation after tail amputation. Transverse sections of 3-day regenerating tails that had been injected with PBS (**e**), anti-GFP antibody (**f**) or anti-AxMLP antibody (**g**) (for details see Extended Data Fig. 9). Sections immunostained for BrdU. * $P < 0.05$, *** $P < 0.0005$ with Student's *t*-test. Scale bar, 200 μm . Yellow circles delineate the spinal cord (top) and notochord/cartilage (bottom).

Δ N-terminal-*AxMlp*-His lacking the target sequence into cultured newt cells (Extended Data Fig. 6a, k). Immunostainings and western blots showed that the morpholinos strongly reduced AxMLP expression but did not affect the expression of Δ N-AxMLP (Extended Data Fig. 6b–j, l–s). No effects were observed with two different negative control morpholinos including a five-mismatch morpholino (Extended Data Fig. 6b–j, l–s).

To knockdown AxMLP *in vivo*, we electroporated *AxMlp* or control morpholinos into the tail epidermis and spinal cord 3 days before amputation. Reduction of protein levels in electroporated cells was confirmed by immunostaining (Extended Data Fig. 7). To test whether exogenously provided AxMLP protein would rescue the knockdown phenotype, the electroporated tails were injected at 1 dpa with purified

AxMLP or inactive flow-through. Blastema length was measured at 1, 3, 6, 10 and 14 dpa, and BrdU incorporation was assayed at 3 dpa (Fig. 4a and Extended Data Fig. 8f, i). Incorporation of BrdU in the *AxMlp* morpholino-electroporated samples was significantly reduced (Fig. 4d and Extended Data Fig. 8a–e). Correspondingly, at 6 dpa, the blastema length of the *AxMlp*-morpholino/flow-through-injected tails was 59% smaller than that of the control morpholino/flow-through-injected ones (specific/flow-through, $546.7 \pm 80.1 \mu\text{m}$; control/flow-through, $1,318 \pm 206 \mu\text{m}$; Fig. 4b, c). In contrast, *AxMlp*-morpholino-electroporated tails injected with purified AxMLP protein showed a 50% rescue in blastema length and 85% rescue of BrdU incorporation (Fig. 4c, d). The partial rescue in blastema length is probably due to the limited amounts of AxMLP provided by a single injection. The specificity of the phenotype was confirmed by implementing the second morpholino and the control five-mismatch morpholino (Extended Data Fig. 8g–j). These results show that knockdown of AxMLP via morpholino results in reduced cell proliferation that can be rescued by provision into the muscle/blastema tissue of exogenous AxMLP protein.

To corroborate the morpholino experiments, we injected the anti-AxMLP blocking antibody into the tail before and during regeneration (Extended Data Fig. 9a), which strongly reduced BrdU incorporation in multiple tissues (Fig. 4e–g and Extended Data Fig. 9b). These results show that *in vivo* knockdown of AxMLP activity by two methods resulted in reduced cell proliferation during early regeneration. To determine if an excess of AxMLP could accelerate regeneration, we performed multiple injections of AxMLP protein before and during early phases of regeneration (Extended Data Fig. 10a). The oversupply of protein resulted in a larger blastema at 4 dpa (Extended Data Fig. 10b, c).

This work represents the first identification of a molecule, AxMLP, by functional expression cloning and *in vivo* testing in appendage regeneration and therefore sets an experimental paradigm for future studies. Previous work indicated that spinal cord neural stem cells accelerate their cell cycle kinetics resulting in increased mitoses between 3 and 4 dpa (ref. 19). Our work indicates that AxMLP is a major factor responsible for the induction of these cell cycle kinetics. How AxMLP is delivered extracellularly, its mode of intracellular signalling and whether orthologues beyond salamanders are associated with regenerative events will be important topics of future investigation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 December 2014; accepted 7 January 2016.

Published online 2 March 2016.

- Wenemoser, D. & Reddien, P. W. Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Dev. Biol.* **344**, 979–991 (2010).
- Knapp, D. et al. Comparative transcriptional profiling of the axolotl limb identifies a tripartite regeneration-specific gene program. *PLoS ONE* **8**, e61352 (2013).
- Tassava, R. A. & Mescher, A. L. The roles of injury, nerves, and the wound epidermis during the initiation of amphibian limb regeneration. *Differentiation* **4**, 23–24 (1975).
- Endo, T., Bryant, S. V. & Gardiner, D. M. A stepwise model system for limb regeneration. *Dev. Biol.* **270**, 135–145 (2004).

- Holtzer, S. The inductive activity of the spinal cord in urodele tail regeneration. *J. Morphol.* **99**, 1–39 (1956).
- Julius, D., MacDermott, A. B., Axel, R. & Jessell, T. M. Molecular characterization of a functional cDNA encoding the serotonin 1c receptor. *Science* **241**, 558–564 (1988).
- Yang, Y. C. et al. Human IL-3 (multi-CSF): identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. *Cell* **47**, 3–10 (1986).
- Tanaka, E. M., Gann, A. A., Gates, P. B. & Brockes, J. P. Newt myotubes reenter the cell cycle by phosphorylation of the retinoblastoma protein. *J. Cell Biol.* **136**, 155–165 (1999).
- Khattak, S. et al. Foamy virus for efficient gene transfer in regeneration studies. *BMC Dev. Biol.* **13**, 17 (2013).
- Khattak, S. et al. Germline transgenic methods for tracking cells and testing gene function during regeneration in the axolotl. *Stem Cell Reports* **1**, 90–103 (2013).
- Schnapp, E. & Tanaka, E. M. Quantitative evaluation of morpholino-mediated protein knockdown of GFP, MSX1, and PAX7 during tail regeneration in *Ambystoma mexicanum*. *Dev. Dyn.* **232**, 162–170 (2005).
- Gruber, C. E. Production of cDNA libraries by electroporation. *Methods Mol. Biol.* **47**, 67–79 (1995).
- Sundaram, M., Cook, H. W. & Byers, D. M. The MARCKS family of phospholipid binding proteins: regulation of phospholipase D and other cellular components. *Biochem. Cell Biol.* **82**, 191–200 (2004).
- Aderem, A. The MARCKS brothers: a family of protein kinase C substrates. *Cell* **71**, 713–716 (1992).
- Sandoval-Guzmán, T. et al. Fundamental differences in dedifferentiation and stem cell recruitment during skeletal muscle regeneration in two salamander species. *Cell Stem Cell* **14**, 174–187 (2014).
- Berg, D. A. et al. Efficient regeneration by activation of neurogenesis in homeostatically quiescent regions of the adult vertebrate brain. *Development* **137**, 4127–4134 (2010).
- Maden, M., Manwell, L. A. & Ormerod, B. K. Proliferation zones in the axolotl brain and regeneration of the telencephalon. *Neural Dev.* **8**, 1 (2013).
- Seykora, J. T., Myat, M. M., Allen, L. A., Ravetch, J. V. & Aderem, A. Molecular determinants of the myristoyl-electrostatic switch of MARCKS. *J. Biol. Chem.* **271**, 18797–18802 (1996).
- Rodrigo Albors, A. et al. Planar cell polarity-mediated induction of neural stem cell expansion during axolotl spinal cord regeneration. *eLife* **4**, e10230 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Tazaki, Y. Taniguchi, I. Wagner, A. Rodrigo-Albors, D. Knapp, P. Murawala, B. Borgonovo and D. Drechsel for technical advice and important discussions; M. Schuez, A. Telzerow and Y. Taniguchi for assistance; B. Gruhl, A. Wagner, S. Mögel for animal care; and J. Currie, T. Sandoval-Guzman and E. Nacu for comments on the manuscript. E.M.T. was supported by a German Federal Ministry of Education and Research (BMBF) Biofutures grant, German Research Foundation (DFG) grant TA274/5-1, European Research Council Advanced Grant, Institutional funding from the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), and the DFG Research Center for Regenerative Therapies Dresden (CRTD) and A.S. by the Swedish Research Council and Cancerfonden.

Author Contributions R.B. performed oocyte injection assay and established expression cloning. T.S. designed and performed expression cloning, *in vitro* cell assays, biochemical experiments, *in vivo* axolotl experiments, analysed experiments and data, and wrote the manuscript. E.M.T. conceived of the project, analysed experiments and data, wrote the manuscript and secured funding. H.W. designed and performed *in vivo* newt experiments, analysed the data and wrote the corresponding parts of the manuscript. A.S. supervised and designed *in vivo* newt experiments, analysed data, edited the manuscript and secured funding.

Author Information The sequence of AxMLP has been deposited in the NCBI GenBank database under accession number KT367888. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M.T. (elly.tanaka@crt-dresden.de) or T.S. (takuji.sugiura@crt-dresden.de).

METHODS

Animals. All animal experiments were performed in accordance with the European Community and local ethics committee guidelines. *Xenopus laevis* were purchased (Nasco) and maintained in our animal facility. *Ambystoma mexicanum* (axolotls) were bred and maintained in our facility, where they were kept at 18 °C in Dresden tap water and fed daily with artemia or fish pellets. Five-to-six centimetre (snout to tail tip) axolotls were used for all the experiments. Animals were anaesthetized for all the surgical process as previously described²⁰. Labelling of connective tissue was achieved by transplanting lateral plate mesoderm from GFP transgenic embryos to normal host embryos as previously described²⁰.

Protein expression by *Xenopus* oocytes. Total RNA was isolated from day-1, day-3 and day-5 limb and tail blastemas with TRIzol reagent (Invitrogen) according to the manufacturer's manual. Total RNA from mature (not regenerating) limb tissues was isolated using the same procedure as blastema samples. Blastema RNAs from the different time points were equivalently pooled as limb-blastema total RNA or tail-blastema total RNA, respectively. mRNA was purified from limb-blastema or tail-blastema total RNA with the Poly (A) Quick mRNA isolation kit (Stratagene). *Xenopus* oocyte preparation and microinjection were performed essentially as previously described^{6,7,21}. Briefly, mature oocytes were defolliculated with collagenase (Sigma). Purified mRNA (5.0 ng) was injected into the selected healthy oocytes after the defolliculation. Eight injected oocytes were cultured together in one well of a 96-well plate (Nunc) for 48 h and the supernatants were harvested for myotube assay (see later).

Functional expression cloning. Clones (110,592) from a 6-day tail blastema library were arrayed into 288 × 384-well plates²². To prepare the 'pools', all the saturated bacterial cultures on one 384-well plate were pooled in one conical tube, and 288 pools were prepared from the library in total. To prepare the 'superpools', 24 pools were combined together in one conical tube, and 12 superpools were prepared from the pools in total (Extended Data Fig. 1a–c). To obtain superpool plasmid, 500 µl of superpool bacteria were cultured in 50 ml LB medium (Extended Data Fig. 1d). To avoid losing low frequency clones in the superpools, the optical density (OD) of each culture was controlled and the cultures were harvested around OD_{600 nm} 0.6. Superpool plasmids were purified with QIAGEN Plasmid Midi Kit (QIAGEN) according to the manufacturer's manual. To reconstitute the whole library, 5 µg of each superpool's plasmids were pooled in one tube before transfection. HEK293FT cells (Invitrogen) were maintained with the standard protocol from Invitrogen. To obtain the superpool supernatants, 8.0 × 10⁵ of HEK293 cells were plated on one well of a 6-well plate (Nunc) and 1 µg of each individual superpool plasmids were transfected into HEK293 cells with Fugene 6 (Roche; Extended Data Fig. 1e) according to the manufacturer's manual. For the first 24 h, the transfected HEK293 cells were kept in the 10% fetal calf serum (FCS) medium. Then the cells were rinsed with FreeStyle 293 expression medium (Gibco) that is a serum-free medium and cultured in the medium at 72 h after transfection. Individually harvested supernatants were concentrated approximately tenfold with a Vivaspin 10,000 MWCO (Sartorius). These concentrated supernatants were tested on A1 myotubes (Extended Data Fig. 1f). It should be noted that given the injury-specific extracellular activity of AxMLP, we infer that the *Xenopus* oocyte and HEK293 cell systems are likely to be in 'wound-epithelium-like' signalling states that permit at least some extracellular release of AxMLP, and that the 6-day regenerating tail blastema cDNA had a sufficient number of AxMLP clones for detection in the expression cloning system. We only detected 1 AxMLP clone among 100,000 clones, and this may reflect the levels of mRNA present at later regeneration time points.

Maintenance of A1 myoblasts, myotube differentiation from A1 myoblasts, myotube purification and subsequent myotube assays were performed essentially as described previously^{8,23,24}. Briefly, concentrated supernatants were individually added into myotube culture medium in a 96-well plate and incubated for 5 days, and BrdU (Sigma; final 10 µg ml⁻¹) was added to the culture media for 18 h before the fixation with 1.5% PFA (Sigma)/PBS. Fixed myotubes were stained with anti-MHC and anti-BrdU antibodies (mouse monoclonal, 4a1025 and Bu20a) conjugated to FITC or rhodamine with DyLight Antibody Labelling Kit (Thermo Scientific) according to the manufacturer's manual. For the quantification of BrdU incorporation activity, the total number of myotubes and BrdU⁺ myotubes were counted by hand under the microscope (Zeiss Axioplan 2). This biological evaluation of the BrdU-incorporation activity on myotubes is called the myotube assay. The high-serum (15% FCS) condition was used as a positive control for myotube assays and the low-serum (0.5% FCS) or serum-free condition was used as a negative control.

For the second screen from superpool number 9, the clones from the 24 384-well plates contained in superpool 9 were divided into smaller sub-pools according to a two-dimensional matrix (Extended Data Fig. 2a). For example, sub-pool A contained pools from plates 193–198 and sub-pool 1 contained pools from plates 193, 199, 205, 211. These sub-pools were cultured to OD_{600 nm} 0.6 before plasmid

preparation. For the third screen from pool number 212, 384 single clones were arrayed by 96-pin plastic replicators (Genetix) on 96-well plates (SARSTEDT) filling 150 µl LB medium per well (Extended Data Fig. 2b; groups A–D). Individual clones on the 96-well plates were statically cultured until they were saturated and 24 clones were pooled together (Extended Data Fig. 2b; sub-pools A1–D4). Plasmids from each pool were purified with QIAprep spin miniprep kit (QIAGEN). For the fourth screen from A1, 24 clones were individually cultured in LB medium and the plasmids were purified with QIAprep spin miniprep kit (QIAGEN). To construct sub-pools, 1 µg of the plasmid from each single clone was pooled according to the diagram in Extended Data Fig. 2c. The process from the transfection into HEK293 cells to myotube assay in the second to the fourth screen was the same as the first screen. To validate transfection efficiency during whole expression cloning, 50 pg of secreted alkaline phosphatase (SEAP)-pCMV-SPORT6 plasmid was co-transfected with the samples as a spike and a portion of the supernatants was assayed by Great EscAPE SEAP Chemiluminescence Kit 2.0 (Clontech). The luminescence of the supernatants was measured by GENiosPro Microplate Reader (TECAN). We confirmed that there was no significant difference of transfection efficiency during the expression cloning (data not shown). There were no cell line misidentification and cross-contamination in the experiments. We used a single mammalian cell line (HEK293FT cells) provided by Invitrogen and single amphibian cell line (newt A1 cells) in the experiments. These two cell lines have totally different morphologies and are cultured under mutually incompatible culture conditions. The growth of both cells were carefully monitored during the experiments and cells samples were constantly stained with Hoechst 33342 (Sigma, final 0.5 µg ml⁻¹) to test mycoplasma contamination.

Plasmid construction. Human and mouse *MLP* cDNA clones were purchased from OriGene Technologies (clone ID: human, SC112373; mouse, MC208965). Zebrafish and *Xenopus mlp* cDNA clones were purchased from Source BioScience (clone ID: zebrafish, 6795591; *Xenopus*, 8330180). All oligonucleotide sequences and the restriction enzyme sites using for cloning are shown in Supplementary Table 2. Since the backbone vector of the cDNA library is pCMV-SPORT6, we subcloned following genes into pCMV-SPORT6 vector (Invitrogen) or pCMV-SPORT6-3C-His vector. PCR-amplified fragments with the oligonucleotides numbers 1 and 2 from pSEAP2-Basic (Clontech) were subcloned into pCMV-SPORT6. The oligonucleotides numbers 3 and 4 were attached to pCMV-SPORT6 to generate a backbone vector, pCMV-SPORT6-3C-His (Extended Data Fig. 3c, bottom left). The *AxMlp* open reading frame (ORF) was amplified by PCR with the oligonucleotides numbers 5 and 6 from the original *AxMlp* clone (*BL212a101*) and subcloned in the pCMV-SPORT6-3C-His vector (Extended Data Fig. 3c, top left). N-terminal deletion *AxMlp* was amplified by PCR with the oligonucleotides numbers 7 and 8 from the original *AxMlp* clone (*BL212a101*) and subcloned in the pCMV-SPORT6-3C-His vector (Extended Data Fig. 6a, bottom). Human, mouse, zebrafish and *Xenopus MLP* ORFs were amplified from purchased cDNA clones with specific primers (for human, oligonucleotides numbers 9, 10; mouse, oligonucleotides numbers 9, 11; zebrafish, oligonucleotides numbers 12, 13; *Xenopus*, oligonucleotides numbers 14, 15, respectively), and were subcloned in the pCMV-SPORT6-3C-His vector. The oligonucleotides numbers 16 and 17 were inserted into the pEGFP-N1 plasmid (Clontech) to generate a backbone vector, pEGFP-N1-3C (Extended Data Fig. 3c, bottom right). The *AxMlp* ORF was amplified by PCR with the oligonucleotides numbers 18 and 19 from the original *AxMlp* clone (*BL212a101*) and subcloned into pEGFP-N1-3C (Extended Data Fig. 3c, top right). Newt *Mlp* ORF was amplified by PCR from newt limb blastema cDNA with the oligonucleotides numbers 28 and 29 and the PCR fragments were subcloned in the pCMV-SPORT6-3C-His vector. These constructs were confirmed by sequencing.

Assessment of AxMLP extracellular secretion. For measuring the GFP intensity of supernatants, 8.0 × 10⁵ of HEK293 cells were plated on 6-well plates and 1 µg of *AxMLP*-3C-pEGFP-N1 plasmid or 1 µg of empty-pEGFP-N1 plasmid were transfected into HEK293 cells. The supernatants were harvested at 24 h post-transfection (hpt), 48 hpt and 72 hpt and concentrated with Vivaspin 10,000 MWCO (Sartorius) individually. The fluorescence intensity was measured using a GENiosPro Microplate Reader (TECAN). To determine the percentage of GFP⁺ cells in the culture, the transfected cells were detached with Trypsin-EDTA (Gibco, final 0.05%)/PBS from the well, then spread on improved Neubauer chamber. The number of GFP⁺ cells and total cells in the grids were counted by hand and the percentage was calculated. Time-lapse imaging was performed under Axiovert 200M (Zeiss) with humidity, temperature and CO₂ control chamber. Images were taken every 30 min from 5 to 72 hpt.

Antibody blocking. For the antibody-based blocking assay *in vitro*, 1 µg of *AxMlp*-3C-His-pCMV-SPORT6 plasmid or empty-pCMV-SPORT6-3C-His plasmid was transfected into HEK293 cells with Fugene 6 (Roche). The supernatants were harvested at 72 hpt and concentrated. Ten micrograms of AxMLP-3C-His protein

(22.7 kDa) were treated with 70 µg or 350 µg of anti-AxMLP polyclonal antibody (see later) or anti-GFP polyclonal antibody (MPI-CBG antibody facility) as a negative control, respectively, at room temperature for 2 h. These antibody-treated supernatants were used for the myotube assay.

For the *in vivo* antibody blocking assay, anti-full-length AxMLP polyclonal antibody (see later), anti-GFP polyclonal antibody (MPI-CBG antibody facility) or PBS as a negative control were injected into mature (not regenerating) tail as the first injection (3 h before amputation) and injected into the blastema as the second injection (12 h post-amputation) and as the third injection (1 day post-amputation) (Extended Data Fig. 9a). These samples were co-injected with tetramethylrhodamine dextran MW 70,000 (Molecular Probes; final 2.5 mg ml⁻¹) as a tracer. The injection efficiency was confirmed based on the intensity of the rhodamine under the fluorescence dissecting microscope (SZX 16, OLYMPUS). No animals were excluded from the analysis. In each injection 500 ng, then, in total 1.5 µg antibody or equivalent volume of PBS were injected. Injected animals were kept in clean tap water for 3 days at room temperature. The animals were injected intraperitoneally with 30 µl of 2.5 mg ml⁻¹ BrdU (Sigma) 4 h before collecting the tails. The injected blastemas were fixed, embedded, cryosectioned and immunostained as described later. For the imaging, the tiled images of the entire cross-section of the tails taken on a Zeiss Observer.Z1 (Zeiss) were then stitched by Axiovision software or Zen 2 (Zeiss). For the quantification at least a total of 1,000 cells per one animal were counted from four different animals in each condition (PBS, anti-GFP antibody or anti-AxMLP injection, respectively), and the marker-positive nuclei (BrdU⁺, PAX7⁺, MEF2⁺ or Hoechst⁺) on the sections were counted by hand. The cells in spinal cord, epidermis and cartilage/notochord were separately counted based on morphology. The label "Other tissues" in Fig. 2d, contained mainly mesenchymal cells and endothelial cells and was calculated by the subtraction from the total number to the number of all the other specific cell types.

AxMLP purification. For His-tagged AxMLP purification, AxMLP-3C-His-pCMV-SPORT6 plasmid was transfected into HEK293 cells and the supernatant was harvested at 72 hpt. His-tagged protein in the supernatant was purified in native conditions on a 1 ml HisTrap HP column (GE Healthcare) using FreeStyle 293 expression medium including 500 mM imidazole step elution. The eluate (purified AxMLP) and depleted media (flow-through) were concentrated with Vivaspin 10,000 MWCO (sartorius) 40 fold and the final concentration of purified AxMLP was 1.31 µg µl⁻¹. Both concentrated eluate (purified AxMLP) and flow-through fractions were dialysed with Spectra/Por Dialysis Membrane MWCO 6-8000 (Spectrum Laboratories) in AMEM (MEM medium (Gibco) diluted 25% with distilled water) for biological assays. The fractions from the purification were tested by silver staining and western blotting (Extended Data Fig. 3g, h). The washing fraction was concentrated about tenfold to load the same volume as other fractions on 4–20% gradient SDS-PAGE gels (anamed Elektrophoresis). Western blotting and silver staining were performed with a standard protocol. Briefly, the fractions were treated with 2× Sample Buffer including dithiothreitol (DTT; Sigma, final 0.2 M) and boiled at 95 °C for 10 min. The proteins were blotted onto PROTRAN nitrocellulose transfer membrane (Whatman) by TE 77 Semi-Dry Transfer Unit (Amersham). The membrane was blocked with 5% skim milk. Primary antibodies used: mouse anti-His (QIAGEN, 1/5,000), mouse anti-α-tubulin (MPI-CBG antibody facility, DM1A 1/5,000), rabbit anti-AxMLP-full length (1/2,500), rabbit anti-AxMLP-C terminus (1/2,500). Secondary antibodies used: goat anti mouse-HRP (Jackson ImmunoResearch Laboratories, 1/5,000), goat anti rabbit-HRP (Jackson ImmunoResearch Laboratories, 1/5,000). Cell lysates for western blotting were obtained by directly adding 2× Sample Buffer on top of the cultured cells and were boiled at 95 °C for 10 min.

Antibodies and immunohistochemistry. For the preparation of anti-full-length AxMLP polyclonal antibody, a glutathione S-transferase (GST) fusion protein with full-length amino acids of AxMLP was expressed in bacteria and purified by standard methods on GS-trap, glutathione sepharose (GE Healthcare). Purified GST fusion protein as an antigen was used to immunize rabbits (Charles River). Anti-serum was affinity purified using maltose-binding protein fused with full-length AxMLP conjugated to NHS-Sepharose resin (GE Healthcare). To raise C-terminal AxMLP polyclonal antibody, keyhole limpet haemocyanin (KLH)-tagged peptides, PPVEPQVEEVAAPAP, was used to immunize rabbits and the affinity-purified polyclonal antibody was provided (Eurogentec). Both anti-full-length and anti-C-terminal AxMLP polyclonal antibodies were tested on the cell lysate from AxMLP-transfected HEK293 cells (Extended Data Fig. 3f).

Limb blastema and tail blastema preparations for sectioning were produced essentially as previously described²⁵. Briefly, limb blastemas amputated at the wrist level were collected from the level of the shoulder, and tail blastemas amputated at the 12th myotome from the cloaca were collected at the 10th myotome of the regenerating tail. These limb and tail blastemas were immunostained as previously described^{15,25,26}. Briefly, the samples were fixed with MEMFA fixative at 4 °C

overnight, and were rinsed with PBS several times. The buffer was replaced from PBS to 10%, 20% and 30% sucrose (Sigma)/PBS, then the samples were embedded in Tissue-Tek O.C.T. Compound (Sakura) for cryosection and the tissues were sectioned at 10-µm thickness with Microm HM 560 cryostat (Thermo). Primary antibodies used: mouse anti-BrdU (MPI-CBG antibody facility, Bu20a 1/400), rabbit anti-BrdU (antibodies-online, 1/600), mouse anti-PAX7 (MPI-CBG antibody facility, PAX7 1/450), rabbit anti-MEF2 (Santa Cruz, 1/200), rabbit anti-AxMLP-C terminus (1/200), rabbit anti-GFP (Rockland, 1/400), rabbit anti-FITC (Invitrogen, 1/400), mouse anti-FITC (Jackson ImmunoResearch Laboratories, 1/400), rat anti-MBP (GeneTex, 1/200). The following appropriate fluorophore-conjugated secondary antibodies were used (all in 1/200 dilution): donkey anti-mouse Alexa Fluor (AF) 647 (Molecular Probes), goat anti-mouse AF 647 (Jackson ImmunoResearch Laboratories), donkey anti-mouse AF 488 (Molecular Probes), goat anti-rabbit AF 647 (Jackson ImmunoResearch Laboratories), donkey anti-rabbit AF 488 (Molecular Probes), donkey anti-rat AF 488 (Molecular Probes). The cell nuclei were stained with Hoechst 33342 (Sigma, final 0.5 µg ml⁻¹). Imaging for the stained sections was performed with Zeiss Observer.Z1 (Zeiss) controlled by Axiovision software or Zen2 (Zeiss).

qRT-PCR. Total RNA preparation, reverse transcription and qRT-PCR were essentially described in the previous work². Briefly, three biological replicas were prepared for each time point and they were technically independent in all the processes (tissue collection, RNA preparation, cDNA synthesis and qRT-PCR). Eight to approximately ten limb or tail blastemas from one time point were collected in one tube and homogenized by POLYTRON PT1600E (KINEMATICA). Total RNA was purified with RNeasy Mini or Midi Kit (QIAGEN) according to the manufacturer's manual. cDNA was synthesized from 300 ng of total RNA using SuperScript III First-Strand Synthesis System (Invitrogen) and qRT-PCR was performed with Power SYBR Green Master Mix (Invitrogen) in total volume of 12 µl with the final primer concentration of 300 nM on the LightCycler 480 (Roche). To obtain the values of fold change for each time point, the relative concentration of the PCR products was calculated by the standard curve method. To obtain the standard curves of the limb time course or the tail time course respectively, the dilution series (1/4, 1/16, 1/64, 1/256 and 1/1,024) were made from the mixture of cDNAs that were equivalently collected from the cDNA samples in all the different time points. These dilution series were used as the template for the PCR and the relative concentrations were calculated by LightCycler 480 Software (Roche) based on the standard curves. The concentration of AxMlp was normalized with that of Rpl4 (large ribosomal protein 4). Primers used for PCR were showed in Supplementary Table 2 (for AxMlp, oligonucleotides numbers 20, 21; for Rpl4, oligonucleotides numbers 22, 23). The raw values of qPCR data are shown in Supplementary Table 1.

Protein injection into axolotl tail and limb. The dialysed protein samples: purified AxMLP, depleted media (flow-through) (see earlier) or PBS as a negative control were injected into mature (not regenerating) tails with a pressure injector, PV830 Pneumatic Picopump (World Precision Instruments). These protein samples were co-injected with tetramethylrhodamine dextran MW 70,000 (Molecular Probes, final 2.5 mg ml⁻¹) as a tracer. A glass capillary (Harvard Apparatus) for the injection was pulled with P-97 Micropipette Puller (Sutter Instrument) and sharpened manually (external tip diameter: 30 µm). The injection efficiency was confirmed based on the intensity of the rhodamine under the fluorescence dissecting microscope (SZX 16, OLYMPUS). No animals were excluded from the analysis. In total, 270 ng of purified AxMLP or equivalent volume of controls were injected into one side of the tail. Injected animals were kept in clean tap water for 3 days at room temperature. The animals were injected intraperitoneally with 30 µl of 2.5 mg ml⁻¹ BrdU (Sigma) 4 h before collecting the tails (Fig. 2a). The injected part of the tails was identified by rhodamine-positive myotomes and these tails were fixed, embedded, cryosectioned and immunostained as described earlier. For the quantification, the tile images of whole cross-sections of the tails from Zeiss Observer.Z1 (Zeiss) were stitched by Axiovision software (Zeiss). Three sections from five different animals in each condition (PBS, flow-through or purified AxMLP injection, respectively) were taken, and the marker-positive nuclei (BrdU⁺, PAX7⁺, MEF2⁺ or Hoechst⁺) on the sections were counted by hand. The cells in spinal cord, epidermis and notochord were separately counted based on morphology. The label "Other tissues", contained mainly mesenchymal cells and endothelial cells, and was calculated by the subtraction from the total number to the number of all the other specific cell types.

For the protein injection into the limb, the procedure was essentially as described earlier. Purified AxMLP protein was injected into the mature (not regenerating) right lower limbs at the centre between the elbow and the wrist. The control samples (flow-through fraction or PBS) were injected into the left limbs of the same animal that were injected with purified AxMLP on their right limbs. In total 2.0 µg purified AxMLP or equivalent volume of controls were injected into

the limbs. The animals were injected intraperitoneally with 30 µl of 2.5 mg ml⁻¹ BrdU (Sigma) 12 h before collecting the limbs (Extended Data Fig. 4e). For the quantification, at least a total of 1,000 cells per one animal were counted from four different animals in each condition (PBS, flow-through or purified AxMLP injection, respectively), and the marker-positive nuclei (BrdU⁺, PAX7⁺, MEF2⁺, MBP⁺ or Hoechst⁺) on the sections were counted by hand. The cells in epidermis and bone/perichondrium were separately counted with their morphology. The label "Other tissues", contained mainly mesenchymal cells and endothelial cells, and was calculated by the subtraction from the total number to the number of all the other specific cell types.

For the acceleration experiment, purified AxMLP, flow-through or PBS as a negative control were injected into mature (not regenerating) tail as the first (3 days before amputation) injection and as the second (1 day before amputation) injection and injected into the blastema as the third (2 days post-amputation) injection (Extended Data Fig. 10a). These samples were co-injected with tetramethylrhodamine dextran MW 70,000 (Molecular Probes, final 2.5 mg ml⁻¹) as a tracer. The injection efficiency was confirmed based on the intensity of the rhodamine under the fluorescence dissecting microscope (SZX 16, OLYMPUS). No animals were excluded from the analysis. The samples were injected into both side of the tail and in each injection, 600 ng, then, in total 1.8 µg protein or equivalent volume of controls were injected. Injected animals were kept in clean tap water for 4 days at room temperature. The length of the blastema was measured from the amputation plane to the tip at the spinal cord level at 4 dpa based on the stereoscope images (SZX 16, OLYMPUS).

Morpholino electroporation. In vitro assay. A1 myoblasts were transfected with original clone *BL212a101*, *AxMLP-3C-His*, *ΔN-AxMLP-3C-His* or empty pCMV-SPORT6-3C-His plasmids and co-transfected with AxMLP-specific morpholinos (Gene Tools; Supplementary Table 2: oligonucleotides numbers 24, 26) or control morpholinos (Gene Tools; Supplementary Table 2: oligonucleotides numbers 25, 27) using Microporator (Digital Bio) according to the manufacturer's manual with some modifications. All morpholinos were modified with FITC at the 3' end. A1 myoblasts were re-suspended in 1 × Steinberg solution at a density of 5.0 × 10⁶ cells per ml followed by incubation of 10 µl cell suspension with 0.5 µg of plasmid and 1 µl of the morpholino (final 100 µM in the incubation). Electroporation was performed at 1,000 V, 35 mS pulse length and 3 pulses and the electroporated cells were spread in 10% FCS AMEM media²⁴ on a 24-well plate (Nunc), immediately after the electroporation. The culture medium was replaced by new media at 24 h post-electroporation and the cells were kept in culture at 72 h post-electroporation. The electroporated cells were fixed with 1.5% PFA/PBS, and the cell lysates were prepared for western blotting. Primary antibodies used for immunostaining: mouse anti-His (QIAGEN, 1/200), mouse anti-FITC (Jackson ImmunoResearch Laboratories, 1/400), rabbit anti-FITC (Invitrogen, 1/400), rabbit anti-AxMLP-full length (1/1,000). Secondary antibodies used for immunostaining (all in 1/250 dilution): goat anti-mouse Cy3 (Jackson ImmunoResearch Laboratories), goat anti-mouse AF488 (Jackson ImmunoResearch Laboratories), donkey anti-rabbit AF 488 (Molecular Probes), goat anti-rabbit Cy3 (Jackson ImmunoResearch Laboratories). Images of the stained cells were taken with Zeiss Observer.Z1 (Zeiss) controlled by Axiovision software (Zeiss).

In vivo assay with rescue protein injection. Electroporation to the spinal cord was performed as previously described with some modifications²⁷. To deliver morpholino into the spinal cord and both sides of the tail epidermis, the tail required electroporation twice with NEPA 21 electroporator (Nepa Gene). The first electroporation was for the spinal cord and one side (left) of the epidermis, and the second electroporation was for the other side (right) of the epidermis. 1.5 µl of morpholino (1.0 mM) was loaded onto a small piece of tissue paper on the left side of the epidermis. Approximately 3 µl of morpholino (1.0 mM) was injected into the spinal cord and immediately electroporated (first electroporation). Sequentially, 1.5 µl of morpholino (1.0 mM) was loaded onto a small piece of tissue paper on the right side of the epidermis and electroporated (second electroporation). The first electroporation conditions: poring pulse, 70 V, 5.0 mS pulse length and 1 pulse; transfer pulse, 55 V, 55 mS pulse length, 5 pulses and 15% decay. The second electroporation conditions: poring pulse, 70 V, 10 mS pulse length and 1 pulse; transfer pulse, 30 V, 30 mS pulse length, 7 pulses and 5% decay. FITC dextran MW 70,000 (Molecular Probes, final 5 mg ml⁻¹) was used as a negative control, since morpholinos were labelled with FITC. The electroporation efficiency in the spinal cord and epidermis was examined based on the intensity of the FITC under the fluorescence dissecting microscope (SZX 16, OLYMPUS). The animals with low FITC intensity were excluded from the next step of the experiments. Three days post-electroporation, the tails were amputated at the level of the maximal morpholino electroporated part. One day post-amputation, a total of 360 ng (180 ng for the spinal cord and 180 ng for blastema) of purified AxMLP or

equivalent volume of control flow-through fraction was injected into the spinal cord and the blastema to rescue the morpholino effect. The length of the blastema was measured from the amputation plane to the tip at the spinal cord level on 1, 3, 6, 10 and 14 dpa based on the stereoscope images (SZX 16, OLYMPUS). To detect BrdU incorporation, the animals were injected intraperitoneally with 30 µl of 2.5 mg ml⁻¹ BrdU (Sigma) 4 h before collecting the tails at 3 dpa. Fixation, embedding, cryosection, staining and imaging were described earlier. For the quantification, 3 cross-sections of the blastema (200–300 µm posterior to the amputation plane) from four different animals in each condition (FITC/flow-through, FITC/purified AxMLP, control morpholino/flow-through, control morpholino/purified AxMLP, AxMLP-specific morpholino/flow-through, AxMLP-specific morpholino/purified AxMLP, respectively) were taken, and the marker-positive nuclei (BrdU⁺, PAX7⁺, MEF2⁺ or Hoechst⁺) on the sections were counted by hand.

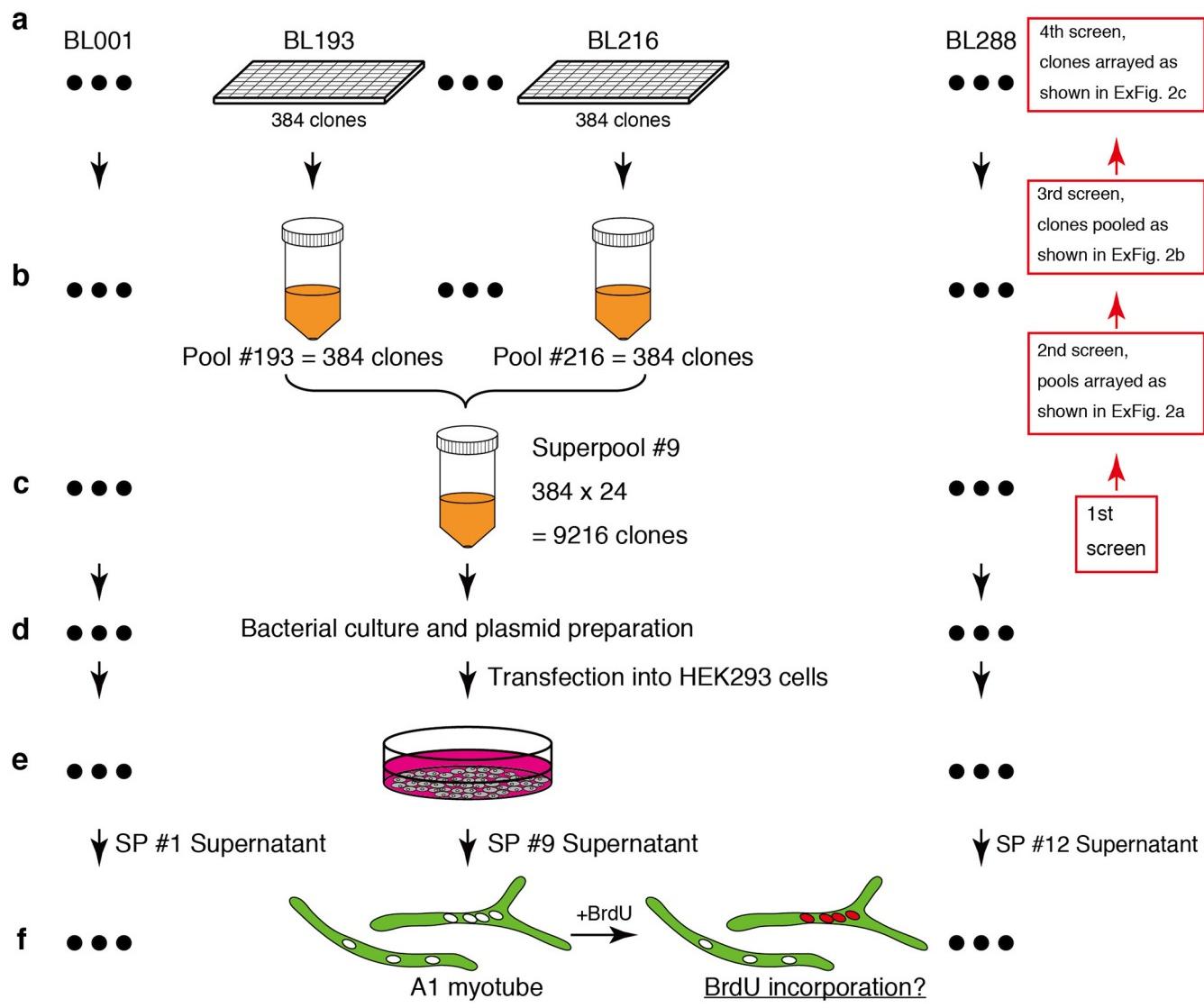
Newt experiments. Animals. Red-spotted newts, *Notophthalmus viridescens*, were supplied by Charles D. Sullivan Co. Animals were anaesthetized in 0.1% ethyl 3-aminobenzoate methanesulfonate (Sigma) for 15 min. Forelimbs were amputated above the elbow, and the bone and soft tissue were trimmed to produce a flat amputation surface. Animals were left to recover overnight in an aqueous solution of 0.5% sulfamerazine (Sigma). At specified time points, the uninjured or regenerating limbs were collected. All surgical procedures were performed according to the European Community and local ethics committee guidelines.

Protein injection and cell cycle assays in newt limbs. The general condition in the newt experiments: 2 µl of 5 mg ml⁻¹ purified AxMLP protein or equivalent volume of flow-through (AxMLP depleted fraction) was injected into the newt limbs. For EdU labelling, animals were injected intraperitoneally with 50–100 µl of 1 mg ml⁻¹ EdU. To investigate the effect of AxMLP on intact newt limbs, purified AxMLP or flow-through was injected into the uninjured limb twice at day 1 and day 3. EdU was administered daily from day 1 to day 5 (Fig. 3a, top). To investigate the effect of AxMLP on regenerating limbs, purified AxMLP or flow-through was injected into the regenerating limbs at 7 and 10 dpa (Fig. 3b, top). EdU was administered daily from 8 to 13 dpa. For labelling myofibre progeny, a *H2B-YFP* reporter construct was introduced into myofibres before amputation as previously described¹⁵ (Fig. 3c, top). Cell cycle re-entry was quantified by EdU incorporation in the YFP⁺ myofibre progeny at 13 dpa.

Immunohistochemistry. Frozen sections (5–10 µm) were thawed at room temperature and fixed in 4% formaldehyde for 5 min. Sections were blocked with 5% donkey serum and 0.1% Triton-X for 30 min at room temperature. Sections were incubated with anti-GFP (Abcam 6673), anti-PAX7 (DSHB) or anti-MHC (DSHB) overnight at 4 °C and with secondary antibodies for 1 h at room temperature. Antibodies were diluted in blocking buffer and sections were mounted in mounting medium (DakoCytomation) containing 5 µg ml⁻¹ DAPI (Sigma). EdU detection was performed as previously described¹⁵. An LSM 700 Meta laser microscope with LSM 6.0 Image Browser software (Carl Zeiss) was used for confocal analyses. One in every eight sections was selected and labelled. For PAX7⁺ satellite cell counting, three sections were randomly selected and counted. For blastema YFP⁺ cell counting, all the sections in the region from regenerate tip to the bone were counted.

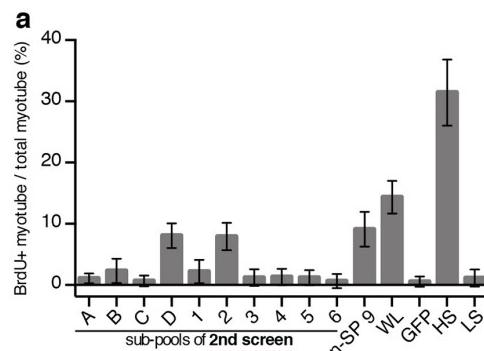
Statistical analysis. Statistical analyses were performed using GraphPad Prism 6.0 (GraphPad Software). Student's *t*-test, parametric, two-tail testing was applied to populations to determine the *P* values indicated in the figures. Significance was considered to have been reached at *P* values from <0.05. No statistical methods were used to predetermine sample size. *In vivo* axolotl experiments were not randomized and no blind tests were applied.

20. Kragl, M. et al. Cells keep a memory of their tissue origin during axolotl limb regeneration. *Nature* **460**, 60–65 (2009).
21. Sive, H. L., Grainger, R. M. & Harland, R. M. *Early Development of Xenopus laevis: A Laboratory Manual* (Cold Spring Harbor Laboratory, 2000).
22. Habermann, B. et al. An *Ambystoma mexicanum* EST sequencing project: analysis of 17,352 expressed sequence tags from embryonic and regenerating blastema cDNA libraries. *Genome Biol.* **5**, R67 (2004).
23. Ferretti, P. & Brookes, J. P. Culture of newt cells from different tissues and their expression of a regeneration-associated antigen. *J. Exp. Zool.* **247**, 77–91 (1988).
24. Lo, D. C., Allen, F. & Brookes, J. P. Reversal of muscle differentiation during urodele limb regeneration. *Proc. Natl. Acad. Sci. USA* **90**, 7230–7234 (1993).
25. Roensch, K., Tazaki, A., Chara, O. & Tanaka, E. M. Progressive specification rather than intercalation of segments during limb regeneration. *Science* **342**, 1375–1379 (2013).
26. Zarzosa, A. et al. Axolotls with an under- or oversupply of neural crest can regulate the sizes of their dorsal root ganglia to normal levels. *Dev. Biol.* **394**, 65–82 (2014).
27. Rodrigo-Albors, A. & Tanaka, E. M. in *Salamanders in Regeneration Research: Methods and Protocols* (eds Kumar, A. & Simon, A.) 115–126 (Springer, 2015).

6-day tail blastema cDNA library: $384 \times 288 = 110,592$ clones

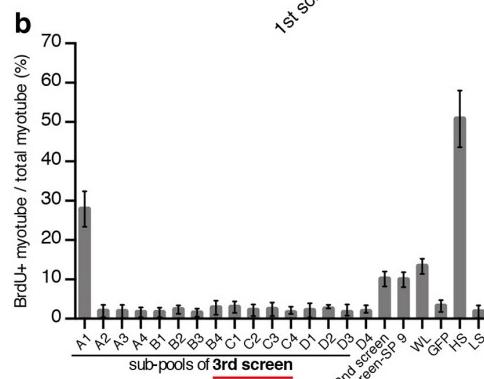
Extended Data Figure 1 | Schematic illustration of the expression cloning approach. **a**, 110,592 clones from a 6-day tail blastema library were arrayed on 288×384 -well plates. **b**, One 384-well plate was pooled into one conical tube and called a ‘pool’. In total, 288 pools were prepared from the library. **c**, Twenty-four pools were combined in one conical tube and called a ‘superpool’ (SP) containing 9,216 clones. In total 12 superpools were prepared. **d**, Bacteria of each superpool was cultured

and plasmid was prepared. **e**, The superpool plasmids were transfected into HEK293 cells. **f**, Individual supernatants were tested on A1 myotubes for cell cycle re-entry activity (myotube assay) (see Fig. 1b). Positive superpools were successively subfractionated and the assay process was repeated back from the positive superpool (first screen) to come to a single clone (fourth screen) (**a–c**, right) (see Extended Data Fig. 2a–c).



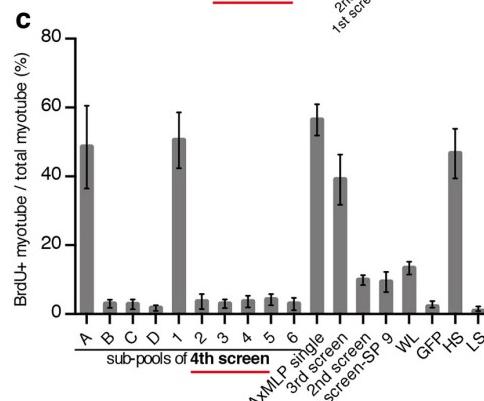
Map of pooled "Pool"

	Pools in Sub-pool 1	Pools in Sub-pool 2	Pools in Sub-pool 3	Pools in Sub-pool 4	Pools in Sub-pool 5	Pools in Sub-pool 6
Pools in Sub-pool A	p193	p194	p195	p196	p197	p198
Pool in Sub-pool B	p199	p200	p201	p202	p203	p204
Pools in Sub-pool C	p205	p206	p207	p208	p209	p210
Sub-pool D	p211	p212	p213	p214	p215	p216



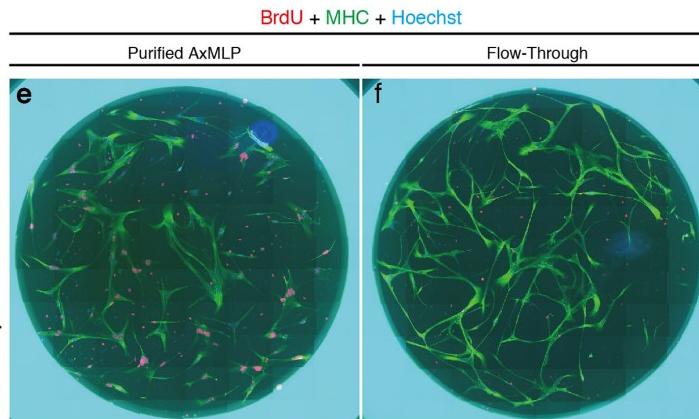
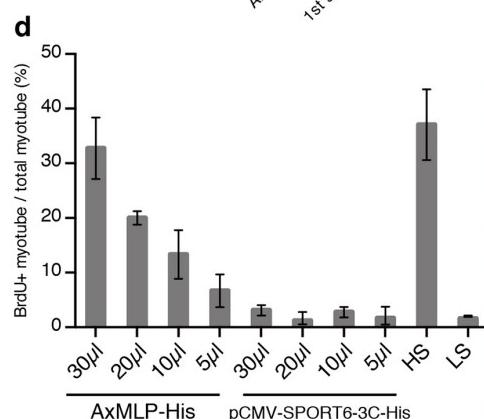
Map of pooled "clones"

BL212 (384 clones)	Group A (96 clones)	Sub-pool A1 (24 clones)
	Group B (96 clones)	Sub-pool A2 (24 clones) Sub-pool A3 (24 clones) Sub-pool A4 (24 clones)
	Group C (96 clones)	Sub-pool B1 (24 clones) Sub-pool B2 (24 clones) Sub-pool B3 (24 clones) Sub-pool B4 (24 clones)
	Group D (96 clones)	Sub-pool C1 (24 clones) Sub-pool C2 (24 clones) Sub-pool C3 (24 clones) Sub-pool C4 (24 clones)
		Sub-pool D1 (24 clones) Sub-pool D2 (24 clones) Sub-pool D3 (24 clones) Sub-pool D4 (24 clones)



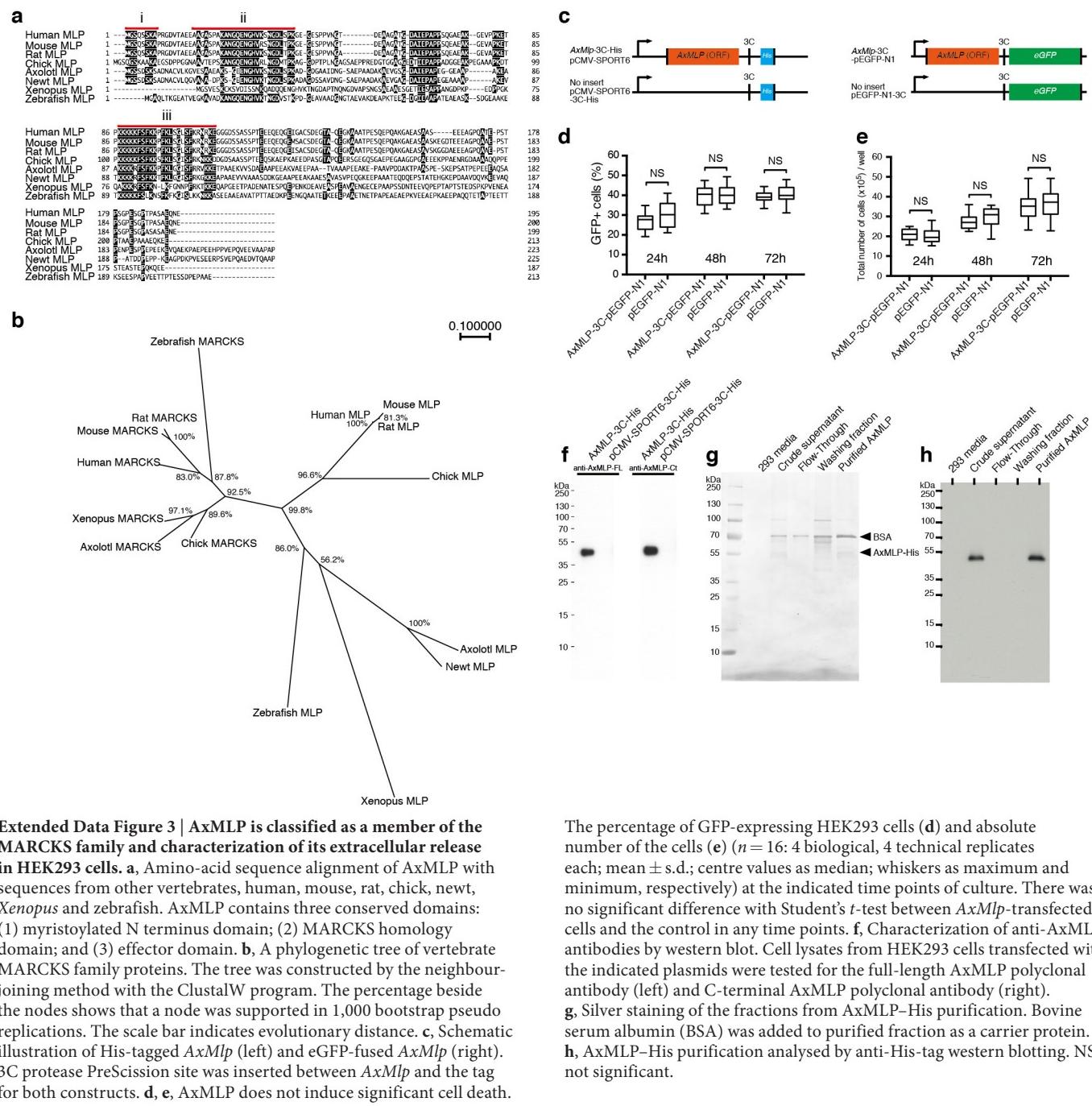
Map of pooled "clones"

	Clones in Sub-pool 1	Clones in Sub-pool 2	Clones in Sub-pool 3	Clones in Sub-pool 4	Clones in Sub-pool 5	Clones in Sub-pool 6
Clones in Sub-pool A	c1	c2	c3	c4	c5	c6
Clones in Sub-pool B	c7	c8	c9	c10	c11	c12
Clones in Sub-pool C	c13	c14	c15	c16	c17	c18
Clones in Sub-pool D	c19	c20	c21	c22	c23	c24



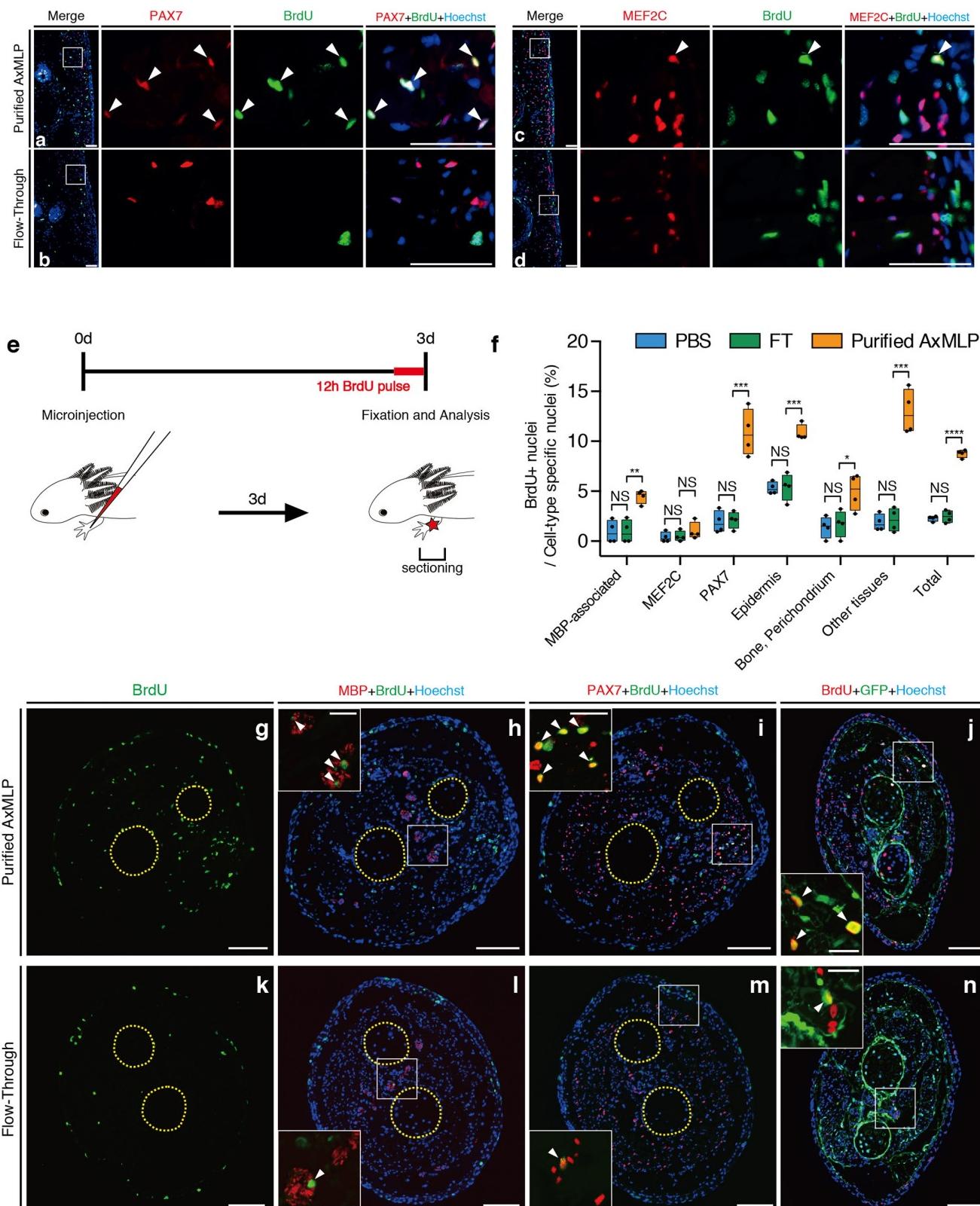
Extended Data Figure 2 | Expression cloning of AxMLP as a myotube cell cycle inducer. **a**, The results of the second-round screen of superpool 9 (see Fig. 1b and Extended Data Fig. 1) and its sub-pooling diagram (right). Sub-pool D and sub-pool 2 showed higher BrdU incorporation activity than the others, identifying pool number 212 as positive ($n = 12$: 4 biological, 3 technical replicates each; mean \pm s.d.). **b**, The result of the third-round screen of pool number 212 from superpool 9 and its sub-pooling diagram (right). Sub-pool A1 showed activity ($n = 6$: 2 biological, 3 technical replicates each; mean \pm s.d.). **c**, Fourth-round screen of SP9 identified a single active clone (c1), AxMlp ($n = 12$: 4 biological, 3 technical replicates each; mean \pm s.d.). The pooling diagram is shown

on the right side. **d**, AxMLP supernatant induces an S-phase response in a dose-dependent manner in the newt myotube assay. Different amounts of AxMLP-containing supernatant (30 µl, 20 µl, 10 µl and 5.0 µl, respectively) were provided to the myotube cell culture medium. The myotube BrdU incorporation correlated with the amount of supernatant provided, whereas pCMV-SPORT6 supernatant did not provoke cell cycle entry at any dose ($n = 6$: 2 biological, 3 technical replicates each; mean \pm s.d.). **e, f**, Newt myotubes treated with purified AxMLP (e) or flow-through (f) were immunostained for BrdU and MHC. More BrdU-incorporated nuclei (red) in myotubes (green) were observed in culture supplied with purified AxMLP compared with flow-through-treated cultures. Scale bar, 1 mm.



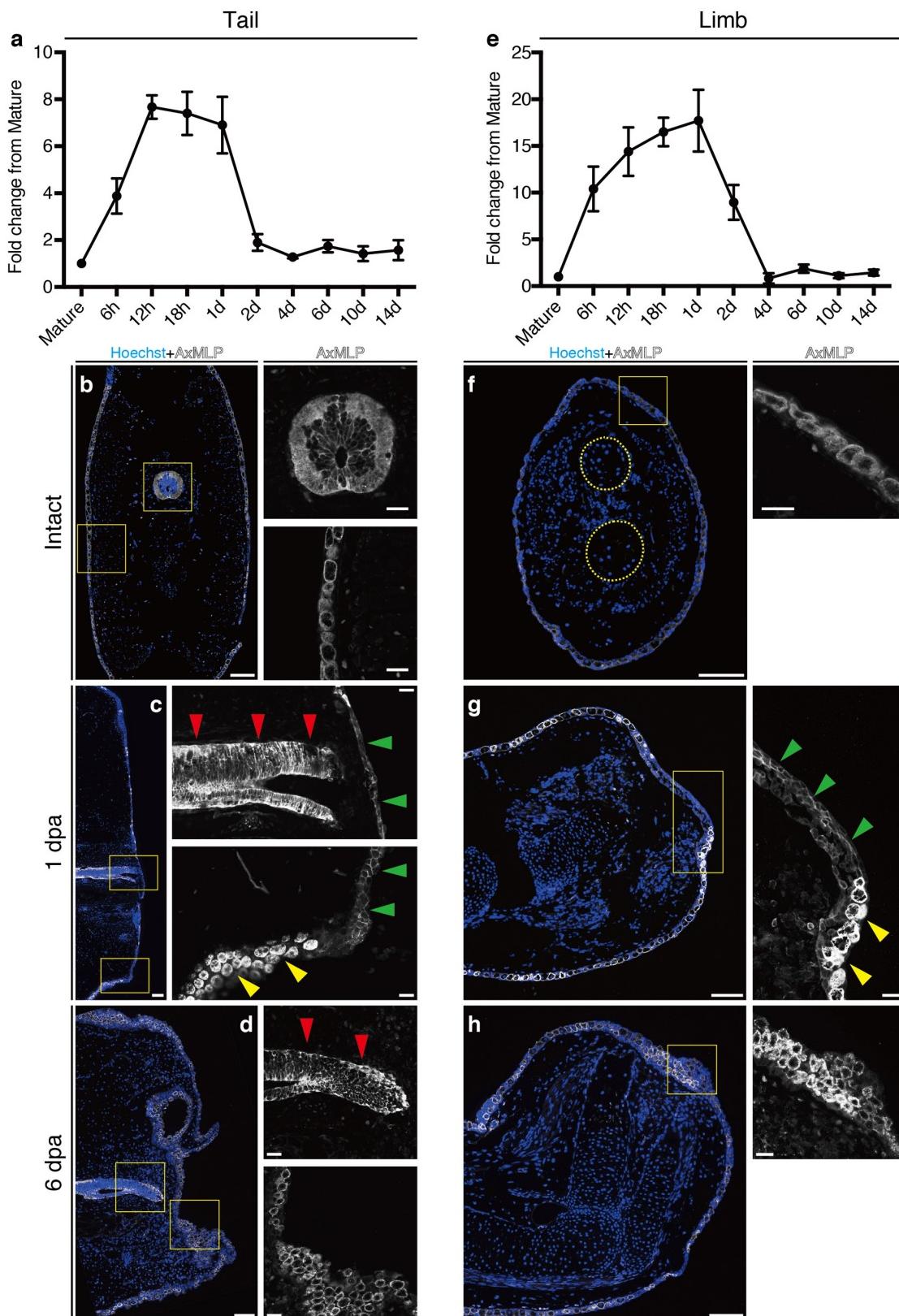
Extended Data Figure 3 | AxMLP is classified as a member of the MARCKS family and characterization of its extracellular release in HEK293 cells. **a**, Amino-acid sequence alignment of AxMLP with sequences from other vertebrates, human, mouse, rat, chick, newt, *Xenopus* and zebrafish. AxMLP contains three conserved domains: (1) myristoylated N terminus domain; (2) MARCKS homology domain; and (3) effector domain. **b**, A phylogenetic tree of vertebrate MARCKS family proteins. The tree was constructed by the neighbour-joining method with the ClustalW program. The percentage beside the nodes shows that a node was supported in 1,000 bootstrap pseudo replications. The scale bar indicates evolutionary distance. **c**, Schematic illustration of His-tagged AxMlp (left) and eGFP-fused AxMlp (right). 3C protease PreScission site was inserted between AxMlp and the tag for both constructs. **d**, **e**, AxMLP does not induce significant cell death.

The percentage of GFP-expressing HEK293 cells (**d**) and absolute number of the cells (**e**) ($n = 16$: 4 biological, 4 technical replicates each; mean \pm s.d.; centre values as median and whiskers as maximum and minimum, respectively) at the indicated time points of culture. There was no significant difference with Student's *t*-test between *AxMlp*-transfected cells and the control in any time points. **f**, Characterization of anti-AxMLP antibodies by western blot. Cell lysates from HEK293 cells transfected with the indicated plasmids were tested for the full-length AxMLP polyclonal antibody (left) and C-terminal AxMLP polyclonal antibody (right). **g**, Silver staining of the fractions from AxMLP-His purification. Bovine serum albumin (BSA) was added to purified fraction as a carrier protein. **h**, AxMLP-His purification analysed by anti-His-tag western blotting. NS, not significant.



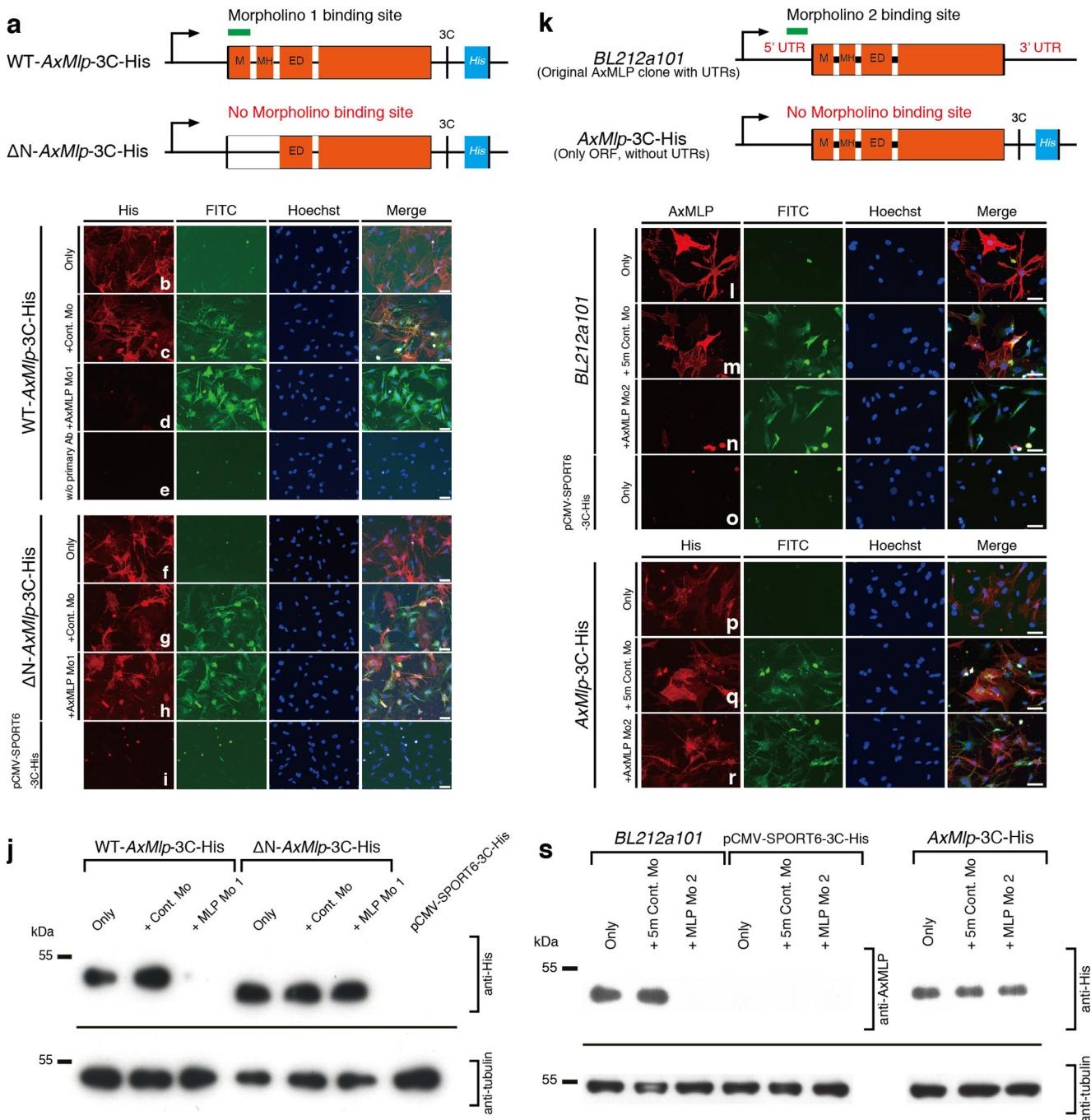
Extended Data Figure 4 | AxMLP is sufficient to induce cell cycle entry in axolotl tail and limb. **a-d**, Sections from AxMLP-injected tails immunostained for BrdU/PAX7 (a, b) and BrdU/MEF2C (c, d) (refers to data in Fig. 2). Scale bars, 100 µm. **e**, Schematic illustration of the protein injection into axolotl limb. **f**, Quantification of BrdU⁺ cells in the limbs injected with PBS, flow-through or purified AxMLP ($n=4$; biological replicates; centre values as median; points represent each sample). **g-n**, Transverse sections from purified AxMLP-injected (g-j) or flow-through-injected limbs (k-n). Scale bars: lower-magnification images, 200 µm; higher-magnification images, 50 µm. Sections were

immunostained for BrdU (g, k), BrdU/myelin basic protein (MBP) (h, l), BrdU/PAX7 (i, m) and BrdU/GFP (j, n). GFP⁺ cells represent connective tissues in lateral plate mesoderm (LPM)-GFP transplanted axolotls. All molecular markers used except MBP had nuclear expression, and therefore allowed one-to-one colocalization of nuclear BrdU with nuclear staining of the marker. Therefore, we refer to the MBP data as ‘MBP-associated’. White boxes highlight the magnified images. Yellow circles indicate two bones in the lower limb. NS, not significant; * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$, **** $P < 0.00005$ with Student’s *t*-test. White arrowheads indicate marker⁺/BrdU⁺ cells.



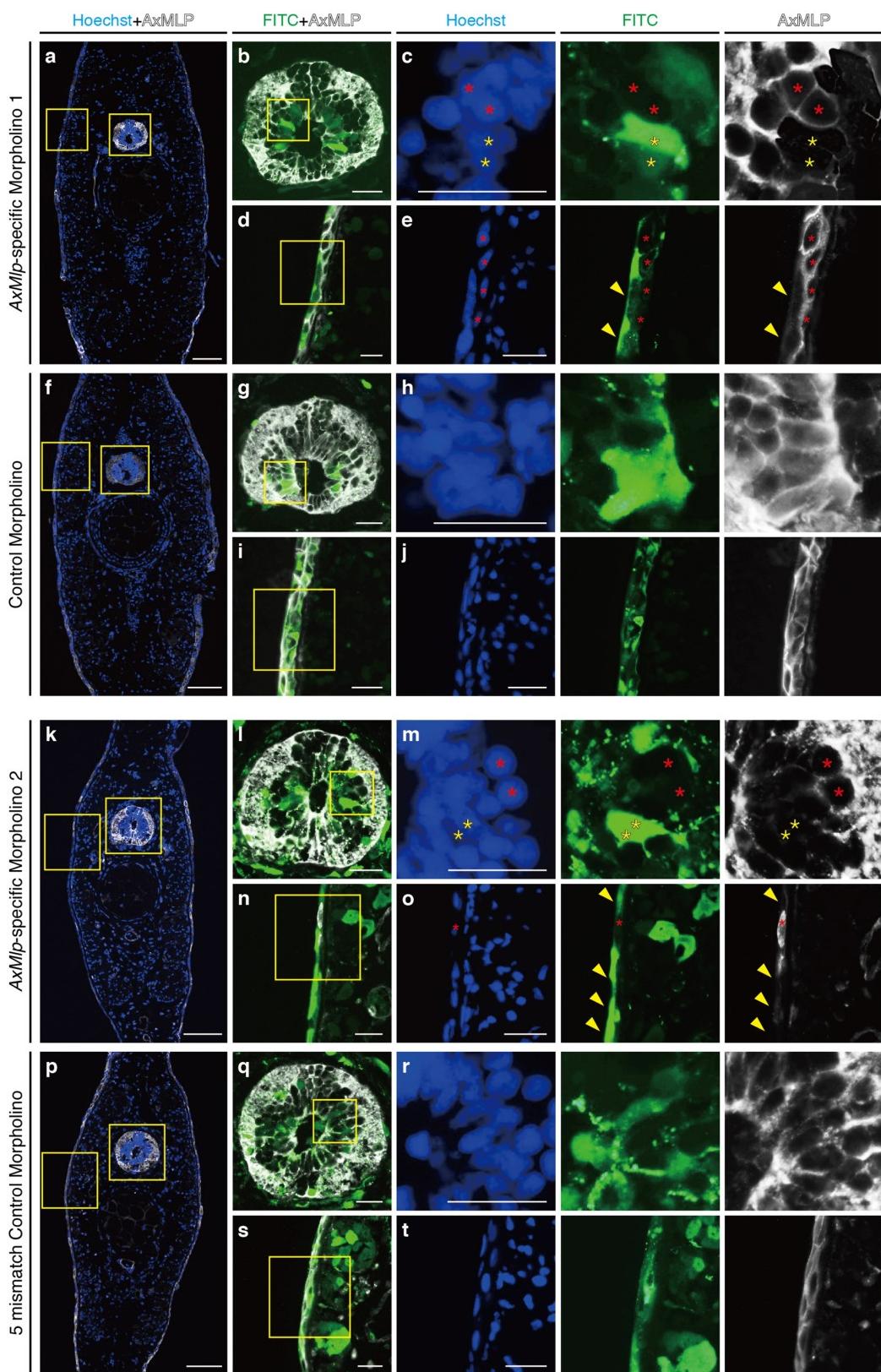
Extended Data Figure 5 | Upregulation of *AxMlp* transcript during early regeneration and alteration of *AxMlp* protein localization in wound epidermis cells. **a, e**, Measurement of *AxMlp* expression by qPCR at the indicated time points during tail (**a**) and limb (**e**) regeneration ($n = 3$: biological replicates; mean \pm s.d.). To obtain the values of fold-change for each time point, the relative concentrations of the PCR products were calculated by the standard curve method. The concentration of *AxMlp* was normalized to that of large ribosomal protein

4 (*Rpl4*). **b–h**, Immunostaining with anti-*AxMlp* antibody (white) on tails (**b–d**) and limbs (**f–h**) of intact (**b, f**; transverse sections), 1 dpa (**c**; sagittal; **g**; horizontal) and 6 dpa (**d**; sagittal; **h**; horizontal) samples. By 6 dpa, the epithelial organization and *AxMlp* expression appeared to be returning to a less tightly adherent, less membrane-associated appearance (**d, h**). Scale bars: left, 200 μ m; right 50 μ m. Red arrowheads indicate spinal cord; green arrowheads indicate wound epidermis; yellow arrowheads indicate normal epidermis; yellow circles indicate two bones in the lower limb.



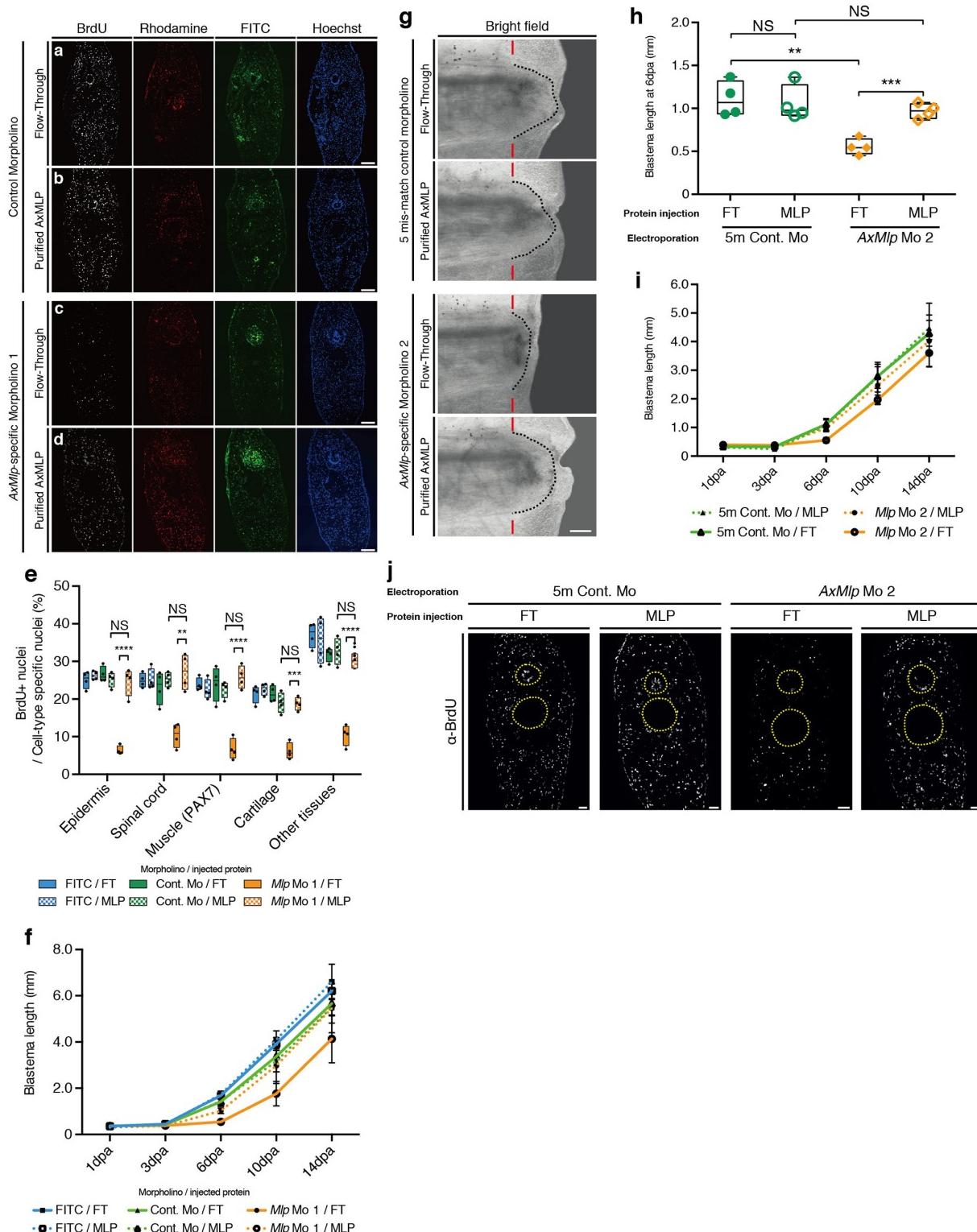
Extended Data Figure 6 | AxMlp morpholinos specifically and efficiently reduce AxMlp translation in cultured cells. **a**, Schematic illustration of wild-type (WT) *AxMlp* (top) and N-terminal deletion *AxMlp* (bottom) constructs used to characterize *AxMlp* morpholino 1. The N-terminally deleted *AxMlp* lacks the morpholino-binding site. Both constructs have a His-tag on their C terminus (**a**). ED, effector domain; M, myristoylated N terminus domain; MH, MARCKS homology domain. **b–e**, Electroporated A1 myoblasts were stained with the indicated markers. **b–e**, Wild-type *AxMlp* plasmid was co-electroporated with the control morpholino (**c**) or the *AxMlp*-specific morpholino 1 (**d**), whereas wild-type *AxMlp* plasmid only (**b**) or wild-type *AxMlp* only without any primary antibody staining were used as negative controls (**e**). **f–h**, Δ N-*AxMlp* plasmid was co-electroporated with the control morpholino (**g**) or the *AxMlp*-specific morpholino 1 (**h**), whereas Δ N-*AxMlp* plasmid only (**f**) or pCMV-SPORT6-3C-His (empty vector) plasmid only served as negative controls (**i**). **j**, Western blotting for the cell lysates from the experiment above. AxMlp morpholino 1 specifically reduced AxMlp protein expression. **k**, Schematic illustration of the

constructs used to characterize *AxMlp* morpholino 2. The original *AxMlp* expression clone from the cDNA library (*BL212a101*; top) was used as it included the 5' untranslated region (UTR) target site for *AxMlp* morpholino 2. The subcloned *AxMlp*-His construct lacks the binding site for *AxMlp* morpholino 2 and was used as the control construct. **l–r**, Electroporated A1 myoblasts were stained with the indicated markers. **l–n**, *BL212a101* plasmid was co-electroporated with the five-mismatch control morpholino (**m**) or the *AxMlp*-specific morpholino 2 (**n**), or *BL212a101* plasmid only (**l**) or pCMV-SPORT6-3C-His (empty vector) plasmid only served as negative controls (**o**). AxMlp was detected using an anti-AxMlp antibody (red), and morpholinos were detected via FITC conjugation (green). **p–r**, *AxMlp*-3C-His plasmid was co-electroporated with the five-mismatch control morpholino (**q**) or the *AxMlp*-specific morpholino 2 (**r**) or *AxMlp*-3C-His plasmid only (**p**). AxMlp was detected using an anti-His-tag antibody (red), and morpholinos were detected via FITC conjugation (green). **s**, Western blotting for the cell lysates from the experiment above. *AxMlp* morpholino 2 specifically reduced AxMlp protein expression. Scale bars, 100 μ m.



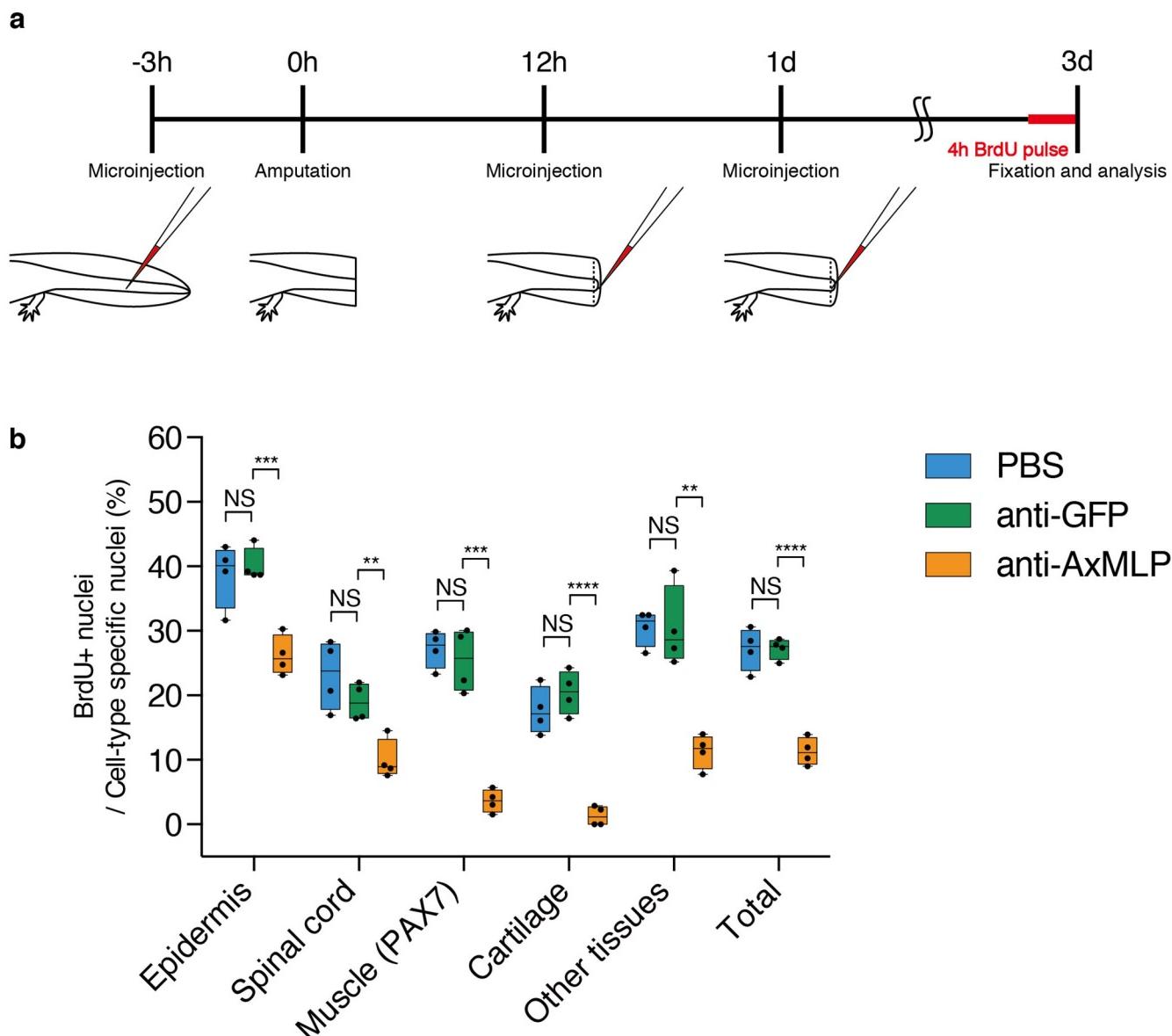
Extended Data Figure 7 | AxMLP morpholinos knockdown endogenous AxMLP *in vivo*. **a–j.** The morpholinos shown in were used in Fig. 4 and Extended Data Figs 6a–j, 8a–f. **k–t.** The morpholinos shown were used in Extended Data Figs 6k–s, 8g–j. Transverse sections from *AxMLP*-specific morpholino 1 (**a–e**) or control morpholino (**f–j**) electroporated tail. **b,** The spinal cord (SC) boxed in **a.** **c,** The higher-magnification images of the spinal cord boxed in **b.** AxMLP expression was detected in morpholino-negative cells (red asterisks), whereas it was reduced in morpholino-positive cells (yellow asterisks). **d,** The epidermis

boxed in **a, e,** AxMLP expression was unaffected in morpholino-negative cells (red asterisks), whereas it was reduced in morpholino-positive cells (yellow arrowheads). In the control morpholino-electroporated tail (**f**) there was no morpholino-specific knockdown phenotype in either spinal cord (**g, h**) or epidermis (**i, j**). The same experiments were performed with *AxMLP*-specific morpholino 2 (**k–o**) and the corresponding five-mismatch control morpholino (**p–t**). **k–t,** The data sets were the same as **a–j.** Scale bars, 200 μ m (**a, f, k, p**); 50 μ m (**b–e, g–j, l–o, q–t**).

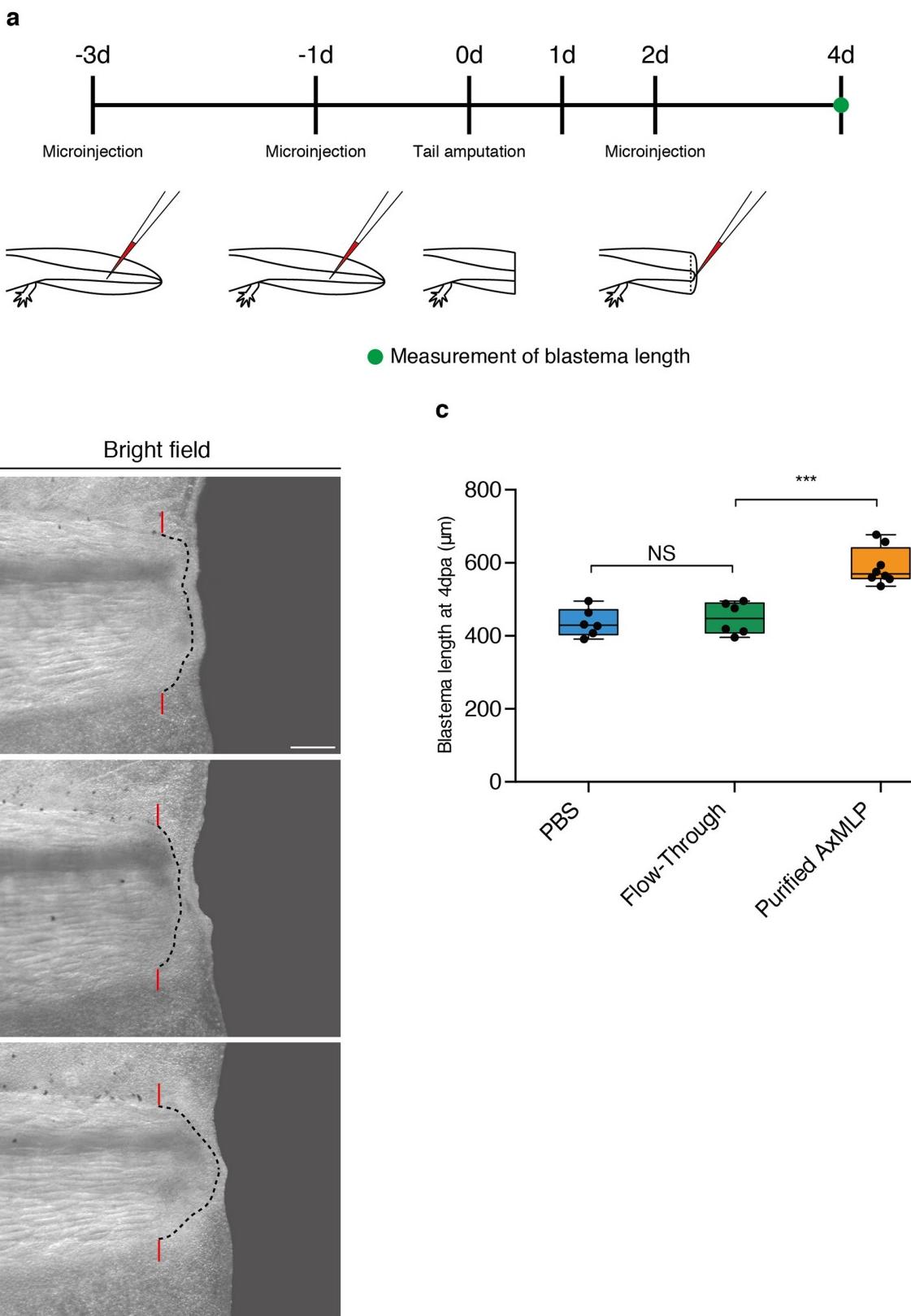


Extended Data Figure 8 | AxMLP is necessary for initial cell proliferation during tail regeneration. **a–d**, Representative transverse sections of the morpholino-electroporated/protein-injected blastemas that were used for quantification of BrdU incorporation in Fig. 4d. Rhodamine was co-injected with the protein samples. **e**, Quantification of BrdU⁺ cells in blastema sections of morpholino-electroporated/protein-injected tails at 3 dpa ($n = 4$: biological replicates; centre values as median; points represent each sample). **f**, The length of the blastema during tail regeneration. The data at 6 dpa were plotted in Fig. 4c. By 14 days the difference in total regenerate length among the samples was not statistically significant. **g–j**, The same experimental scheme (shown in Fig. 4a) as was used for AxMLP morpholino 1 was implemented for a second specific morpholino (AxMLP-specific morpholino 2). **g**, Bright-field images of the morpholino-2-electroporated/

protein-injected tails at 6 dpa. Red bars indicate amputation planes. Dashed lines delineate the shape of the mesenchymal blastema. **h**, Blastema length at 6 dpa ($n = 4$: biological replicates; centre values as median; points represent each sample). **i**, The length of the blastema during tail regeneration. The data at 6 dpa were plotted in **h**. **j**, Transverse sections immunostained for BrdU from morpholino-electroporated/protein-injected tails at 3 dpa. AxMLP-specific morpholino 2 combined with flow-through (FT) injection shows reduction of BrdU incorporation, whereas AxMLP protein injection rescues the phenotype. The corresponding five-mismatch control morpholino does not affect BrdU incorporation. Yellow circles indicate spinal cord (top) and notochord/cartilage (bottom). NS, not significant; ** $P < 0.005$, *** $P < 0.0005$, **** $P < 0.00005$ with Student's *t*-test. Scale bars, 200 μ m (**a–d, j**); 500 μ m (**g**).



Extended Data Figure 9 | Anti-AxMLP antibody significantly blocks BrdU incorporation during tail regeneration. **a**, Schematic illustration of antibody injection into axolotl tail. **b**, Quantification of BrdU⁺ cells in blastema sections of antibody-injected tails at 3 dpa ($n = 4$: biological replicates; centre values as median; points represent each sample). NS, not significant; ** $P < 0.005$, *** $P < 0.0005$, **** $P < 0.00005$ with Student's *t*-test.



Extended Data Figure 10 | Exogenous AxMLP accelerates normal tail regeneration. **a**, Schematic illustration of the protein injection into axolotl tail and blastema. **b**, Bright-field images of the protein-injected tails at 4 dpa. **c**, Blastema length at 4 dpa ($n = 6$: PBS, FT; $n = 8$: AxMLP, biological replicates; centre values as median; points represent each sample). The

blastema from purified AxMLP injected tails significantly increased the regenerate length. Scale bar, 500 μm . Red bars indicate amputation planes; dashed lines delineate the shape of the mesenchymal blastema. NS, not significant; *** $P < 0.0005$ with Student's *t*-test.

A receptor heteromer mediates the male perception of female attractants in plants

Tong Wang^{1,2}, Liang Liang^{1,2}, Yong Xue^{1,2}, Peng-Fei Jia¹, Wei Chen^{1,2}, Meng-Xia Zhang^{1,2}, Ying-Chun Wang¹, Hong-Ju Li¹ & Wei-Cai Yang¹

Sexual reproduction requires recognition between the male and female gametes. In flowering plants, the immobile sperms are delivered to the ovule-enclosed female gametophyte by guided pollen tube growth. Although the female gametophyte-secreted peptides have been identified to be the chemotactic attractant to the pollen tube^{1–3}, the male receptor(s) is still unknown. Here we identify a cell-surface receptor heteromer, MDIS1–MIK, on the pollen tube that perceives female attractant LURE1 in *Arabidopsis thaliana*. MDIS1, MIK1 and MIK2 are plasma-membrane-localized receptor-like kinases with extracellular leucine-rich repeats and an intracellular kinase domain. LURE1 specifically binds the extracellular domains of MDIS1, MIK1 and MIK2, whereas *mdis1* and *mik1 mik2* mutant pollen tubes respond less sensitively to LURE1. Furthermore, LURE1 triggers dimerization of the receptors and activates the kinase activity of MIK1. Importantly, transformation of AtMDIS1 to the sister species *Capsella rubella* can partially break down the reproductive isolation barrier. Our findings reveal a new mechanism of the male perception of the female attracting signals.

Peptides have recently been identified as female attractants, such as *Zea mays* EGG APPRATUS 1 (ZmEA1) in maize, defensin-like peptides LURE1 and LURE2 in *Torenia fournieri* (TfLURE1 and TfLURE2) and AtLURE1 in *A. thaliana*^{1–3}. However, the receptor(s) in the pollen tube perceiving the female attractants is not known. To identify the male receptors, we selected receptor-like kinases (RLKs) preferentially expressed in *Arabidopsis* pollen (tubes)^{4–7} as candidates (Extended Data Fig. 1a). To investigate their function, the kinase-dead dominant negative (DN) forms were expressed in wild-type plants under the pollen-specific *LAT52* (ref. 8) promoter. Micropylar targeting of the RLK^{DN}-expressing pollen tubes was analysed under minimal pollination⁹. Second, we analysed the micropylar targeting of the pollen tubes of the corresponding knockout mutants. Third, we examined possible interactions between these RLKs for potential co-receptors by yeast two-hybrid analysis. Through this combinatory approach, two homologous leucine-rich-repeat RLKs clades, At5g45840 and At4g18640 (previously designated as MRH1 (ref. 10)), and At4g28650 and At4g08850, were identified and designated MALE DISCOVERER1 (MDIS1) and MDIS2, and MDIS1-INTERACTING RECEPTOR LIKE KINASE1 (MIK1) and MIK2, respectively (Extended Data Fig. 1b). *MDIS1*^{DN} pollen tubes exhibit decreased micropylar guidance (Extended Data Fig. 1c–f) and fertilization efficiency (Extended Data Fig. 1g) in the T1 hemizygotes and T3 homozygotes compared to the wild type. The progenies of two single T-DNA insertion lines (*MDIS1*^{DN}-1 and *MDIS1*^{DN}-2) segregate at 2.3:1 and 2.2:1 for the transgenes. During reciprocal crosses, decreased male transmission was observed, but not reduced female transmission (Extended Data Table 1). This result indicates that *MDIS1*^{DN} interferes with the pollen tube guidance. Furthermore, MDIS1 interacts with MIK1 and MIK2 in yeast (Extended Data Fig. 1h). Genomic-fused GUS and green

fluorescent protein (GFP) reporters further confirmed the expression of *MDIS1* and *MDIS2* in pollen tubes and seedlings, and their localization in plasma membrane and endomembrane compartments, respectively (Fig. 1a–d, Extended Data Fig. 2a–c and Supplementary Videos 1 and 2). Corroboratively, *MDIS1*, *MDIS2*, *MIK1*, *MIK2* and the close homologue of *MIK1*, *PXY*, were predominantly expressed in pollen tubes (Extended Data Fig. 2d). *PXY* has been shown to be the receptor of TDIF in vascular development^{11,12} and was detected at low level in pollen and pollen tubes. Genetic results showed that MIK1 may not be the receptor of TDIF¹², but it cannot be excluded that MIK1 might be the receptor of other pistil-expressed CLE peptides. Immunostaining revealed the expression of *MIK1* and *MIK2* in pollen tubes (Fig. 1e–j and Extended Data Fig. 2e, f). These results suggest that they function in pollen tubes.

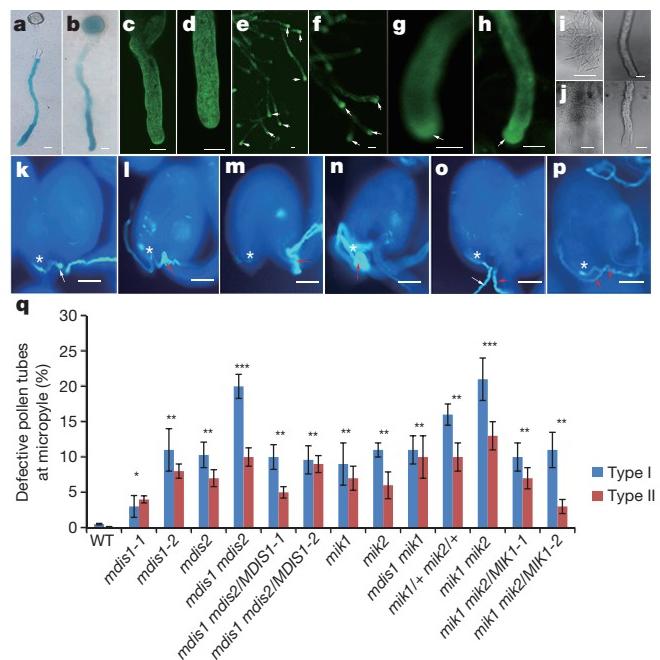


Figure 1 | Expression of MDIS1, MDIS2, MIK1 and MIK2 and their mutant phenotype. a–d, MDIS1-GUS (a), MDIS2-GUS (b), MDIS1-GFP (c) and MDIS2-GFP (d) in pollen tubes. e–j, Wild-type (e–h), *mik1* (i) and *mik2* (j) tubes stained with MIK1 (e, g) or MIK2 (f, h) antibody. Arrows denote tube tips. Scale bars, 5 μm. k–p, Phenotype of wild-type (k) and mutant (l–p, red arrows) pollen tubes at the micropyle (asterisks). Images are representative of 30 images captured. Scale bars, 50 μm. q, Statistical analysis. Error bars, s.e.m. of 3 independent replicates; n = 300 for each sample. *P < 0.05, **P < 0.01, ***P < 0.001 (Student's t-test). -1 and -2 are genetic complementation lines.

¹State Key Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. ²University of Chinese Academy of Sciences, Beijing 100049, China.

To investigate their roles in the pollen tube, knockout mutants *mdis1-2*, *mdis2*, *mik1* and *mik2*, and a knockdown mutant *mdis1-1* were obtained, and *mdis1-2* was used for *mdis1* analysis (Extended Data Fig. 3a, b). During reciprocal crosses with *mdis1-1/+ mdis2-/-* or *mdis1-1/- mdis2-/+*, we observed reduced male transmission and normal female transmission (Extended Data Table 2). Furthermore, the *in vivo* tube length and *in vitro* pollen germination ratio of *mdis1 mdis2*, *mik1 mik2* and *mik1 mik2* were normal (Extended Data Fig. 3c–e). When growing in the wild-type pistils, the wild-type pollen tubes enter the micropyle directly (Fig. 1k). The mutants, however, displayed two major types of defective pollen tube responses to the ovules (Fig. 1l–q), that is, type I is featured by failed pollen tube entry (Fig. 1l–n), type II is featured by one pollen tube failing but another tube entering (Fig. 1o) and occasionally *mdis1* and *mdis1 mdis2* pollen tubes branching at the micropyle (Fig. 1p). The type II phenotype may explain the lack of seed set defect under natural pollination. To confirm this hypothesis, we counted the number of earlier (appeared larger) and later fertilized wild-type ovules by the *mik1 mik2* pollen under limited pollination. The ratio of later to earlier fertilized ovules by the mutant pollen tubes is higher than that by the wild-type pollen tubes (Extended Data Fig. 3f, g), indicating that the fertilization efficiency of mutant pollen tubes is decreased. The *mdis1 mdis2* and *mik1 mik2* double mutations exaggerate the guidance defect, but *mdis1 mik1* did not (Fig. 1q), indicating that *MDIS1/MDIS2* and *MIK1/MIK2* probably act in the same pathway. The full-length genomic sequence of *MDIS1-GFP* and the *MIK1* coding sequence driven by the *LAT52* promoter alleviates the phenotype of *mdis1 mdis2* and *mik1 mik2* to the single mutant (Fig. 1q). These data indicate that both MDIS and MIK have a role in the tube perception of the female signal.

To verify if MDIS1, MDIS2, MIK1 and MIK2 are the receptors of LURE1, we examined the binding of AtLURE1.2 with the purified recombinant ectodomain (ECD) of MDIS1, MDIS2, MIK1 and MIK2. The glutathione S-transferase (GST)-tagged ECD of MDIS1, MIK1 and MIK2 binds His-tagged LURE1.2, but not to the His-tagged defensin-like peptide AtPDF1.2 (ref. 13), as shown by a pull-down assay (Fig. 2a). AtPRK3 (AT3G42880), a leucine-rich repeat RLK highly expressed in the pollen tube¹⁴, does not bind LURE1.2. Consistently, the purified proteins are properly folded demonstrated by mass spectrometry analysis that showed that the disulfide bonds between the two cysteine residues in the amino- or carboxy-terminal capping domains of the purified MDIS1^{ECD}, MIK1^{ECD} and MIK2^{ECD} were properly formed (Extended Data Fig. 4). Furthermore, microscale thermophoresis (MST) analysis showed that LURE1.2 strongly interacts with MDIS1, MIK1 and MIK2, with a dissociation constant (K_d) of $1.76 \pm 0.09 \mu\text{M}$, $672 \pm 42.4 \text{ nM}$ and $464 \pm 13.4 \text{ nM}$, respectively, but MDIS2 and PRK3 exhibit no binding with LURE1.2 (Fig. 2b). The ERECTA protein previously shown to bind TflURE2 at a background affinity ($279 \pm 60 \text{ nM}$) using a microsome and Quartz crystal microbalance method¹⁵, also displayed a background affinity binding to TflURE2 ($94.6 \pm 2.46 \mu\text{M}$) using MST analysis (Extended Data Fig. 5a). The discrepancy in affinity probably resulted from different methods used and it is common that the affinity derived from a cell-based assay is much higher than an *in vitro* protein-based assay; presumably they differ in cellular context and other signalling components. Interestingly, the MIK1 homologue PXY also binds His-LURE1.2, with a dissociation constant of $704 \pm 49.2 \text{ nM}$ in the MST assay. The finding that MDIS2 does not bind LURE1 and the additive phenotype of *mdis1 mdis2* suggest that MDIS2 may bind other unidentified female attractants, as suggested by the partial

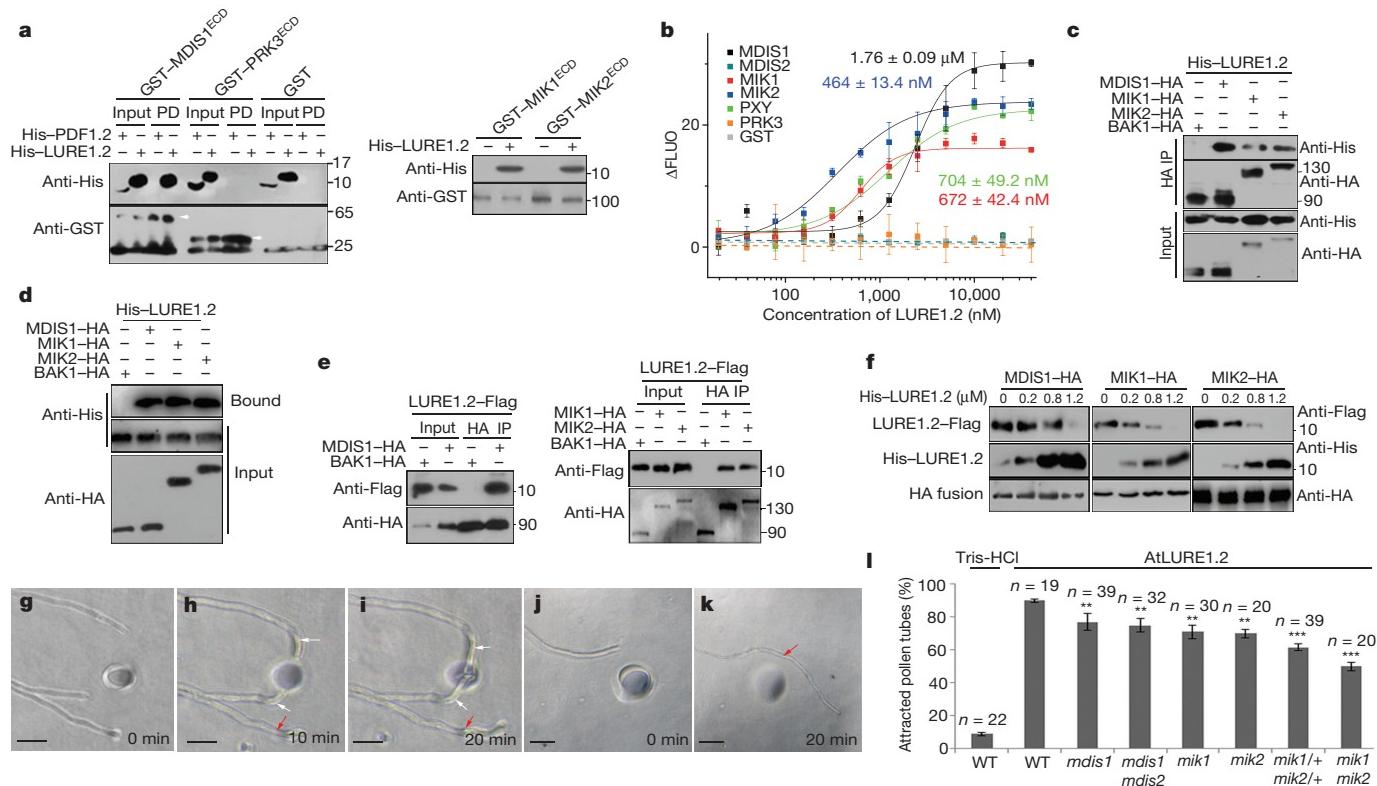


Figure 2 | MDIS1, MIK1 and MIK2 are LURE1 receptors. **a**, Pull-down (PD) assay as indicated. **b**, Binding affinity by MST. Error bars, s.e.m. of 3 independent measurements. ΔFLUO , change in fluorescence. **c**, Interactions between HA fusions and His-LURE1.2 in protoplasts. **d**, His-LURE1.2 binds the protoplasts expressing HA fusions. **e**, Co-IP between HA fusions and LURE1.2-Flag. **f**, Competition between LURE1.2-Flag and His-LURE1.2 to HA fusions. Full blots are shown in

Supplementary Data. **g–l**, Growth of wild-type (**g–i**) and mutant (**j, k**) pollen tubes to the LURE1.2 beads. Red arrows, unattracted; white arrows, attracted. Images are representative of 30 images captured. Scale bars, 20 μm . **m**, Attraction frequency. n , number of pollen tubes scored. Error bars, s.e.m. of 3 independent replicates. ** $P < 0.01$, *** $P < 0.001$ (Student's *t*-test).

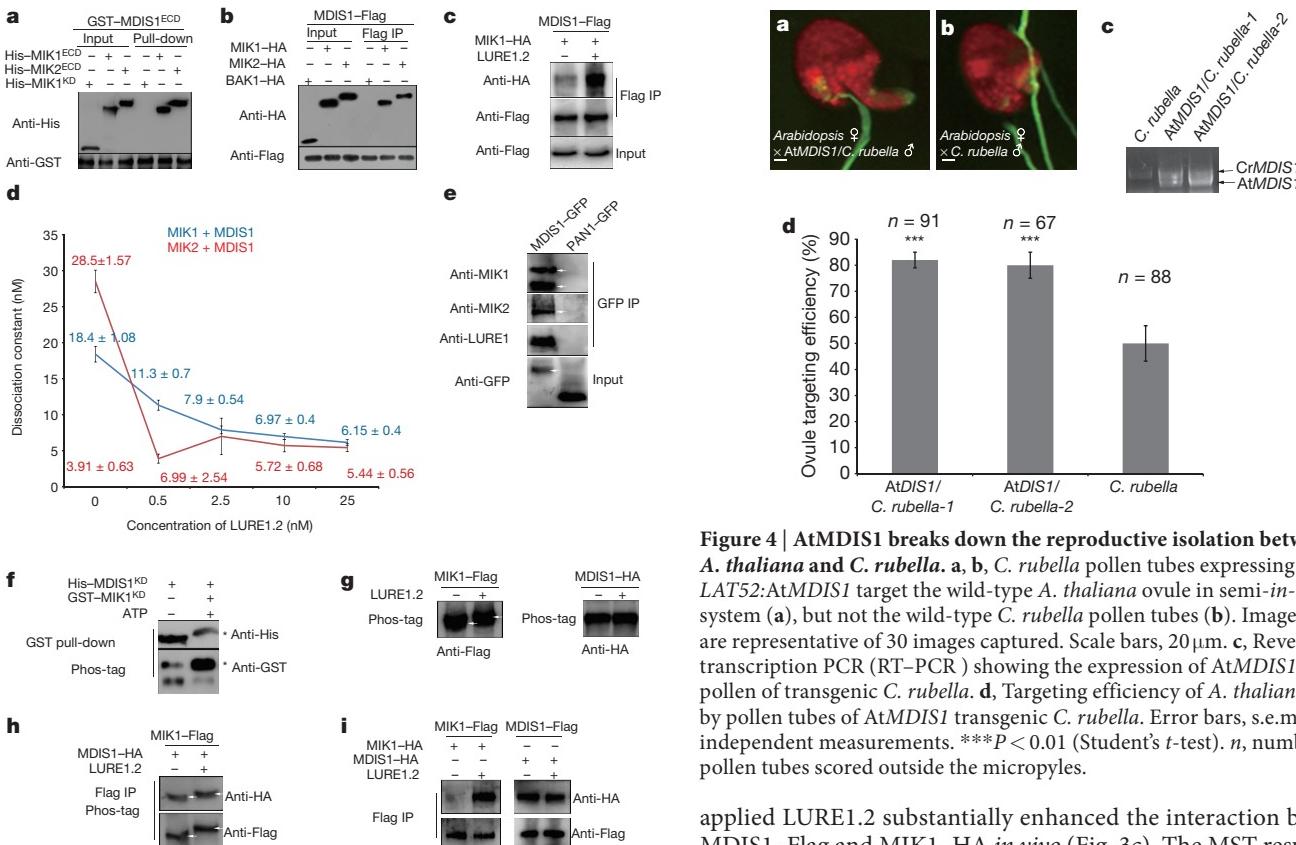


Figure 3 | MDIS1 and MIKs synergistically perceive LURE1. **a**, Pull-down assay as indicated. **b, c**, LURE1 enhances the interaction between MDIS1-Flag and MIK1-HA or MIK2-HA by co-IP in protoplasts. **d**, Affinity between GST-MDIS1^{ECD} and His-MIK1^{ECD} or MIK2^{ECD} in the presence of His-LURE1.2 by MST. Error bars, s.e.m. of 3 independent measurements. **e**, MDIS1-GFP interacts with MIK1, MIK2 and LURE1 in planta. Arrows denote target proteins. **f**, GST-MIK1^{KD} phosphorylated itself and His-MDIS1^{KD}. Asterisks denote phosphorylated proteins. **g**, LURE1.2 induces MIK1-Flag self-phosphorylation. **h**, MIK1-Flag phosphorylates MDIS1-HA and itself after LURE1.2 treatment. **i**, LURE1.2 induces homodimerization of MIK1. Full blots are shown in Supplementary Data.

guidance defect of LURE1-knockdown plants³. We further confirmed the binding of MDIS1, MIK1 and MIK2 to LURE1.2 by co-immunoprecipitation (co-IP) in *Arabidopsis* leaf protoplasts. The haemagglutinin (HA)-tagged full-length MDIS1, MIK1 and MIK2 bind His-LURE1.2, but HA-tagged BRI1-ASSOCIATED KINASE1 (BAK1)¹⁶ does not (Fig. 2c). Furthermore, His-LURE1.2 is associated with the protoplasts expressing MDIS1-HA, MIK1-HA and MIK2-HA (Fig. 2d). The Flag-tagged LURE1.2 purified from *LURE1.2*-overexpressing plants was co-immunoprecipitated by protoplast-expressed MDIS1-, MIK1- and MIK2-HA, respectively (Fig. 2e). The binding of plant-purified LURE1.2-Flag to MDIS1, MIK1 and MIK2 was competitively replaced by an excess of His-LURE1.2, suggesting that the bindings are specific (Fig. 2f). Furthermore, we demonstrated that LURE1.2 triggers endocytosis of MDIS1-GFP in the pollen tube tip (Extended Data Fig. 5b–e). Consistently, the wild-type pollen tubes were attracted to the LURE1.2-embedded beads efficiently, while the mutant tubes show a significantly reduced response to the attractant (Fig. 2g–l), in the semi-in-vitro guidance assay³. The above data showed that MIDS and MIK bind LURE1 both *in vitro* and *in vivo*.

Next, we explored whether MIK1 and MIK2 might work synergistically with MDIS1. Direct interactions between MDIS1^{ECD} and MIK1^{ECD} or MIK2^{ECD} were detected in pull-down and co-IP assays (Fig. 3a, b and Extended Data Fig. 5f, g). Importantly, exogenously

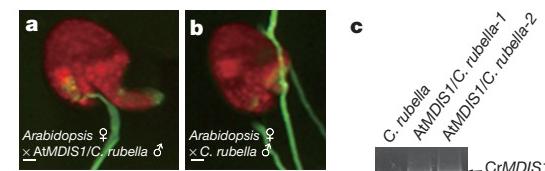


Figure 4 | AtMDIS1 breaks down the reproductive isolation between *A. thaliana* and *C. rubella*. **a, b**, *C. rubella* pollen tubes expressing *LAT52:AtMDIS1* target the wild-type *A. thaliana* ovule in semi-in-vitro system (**a**), but not the wild-type *C. rubella* pollen tubes (**b**). Images are representative of 30 images captured. Scale bars, 20 μm. **c**, Reverse transcription PCR (RT-PCR) showing the expression of *AtMDIS1* in the pollen of transgenic *C. rubella*. **d**, Targeting efficiency of *A. thaliana* ovules by pollen tubes of *AtMDIS1* transgenic *C. rubella*. Error bars, s.e.m. of 3 independent measurements. ***P < 0.01 (Student's t-test). n, number of pollen tubes scored outside the micropyles.

applied LURE1.2 substantially enhanced the interaction between MDIS1-Flag and MIK1-HA *in vivo* (Fig. 3c). The MST result verified that LURE1.2 enhances the interaction between MDIS1^{ECD} and MIK1^{ECD} or MIK2^{ECD} (Fig. 3d). Furthermore, bimolecular fluorescence complementation confirmed that LURE1.2 enhances the interaction between MDIS1 and MIK proteins (Extended Data Fig. 6a–f). An *in planta* co-IP assay with self-pollinated flowers of *MDIS1-GFP* transgenic plants confirmed that the MIK-MDIS1 complex perceives LURE1 (Fig. 3e and Extended Data Fig. 6g, h).

Ligand-induced heterodimerization of co-receptor complex transduces signals by transphosphorylation during pathogen and brassinosteroid perception^{17,18}. We determined whether this is true for MIK and MDIS since MDIS1 and MDIS2 are atypical RLKs¹⁹. Using a Phos-tag mobility shift assay, we found that the kinase domain of MDIS1 (MDIS1^{KD}) was phosphorylated by MIK1^{KD}, which exhibits self-phosphorylation, whereas MDIS1^{KD} shows no self-phosphorylation (Fig. 3f). By mass spectrometry, we found that MDIS1 is phosphorylated by MIK1 at Ser663, and MIK1 is auto-phosphorylated at eight sites (Thr741, Thr742, Thr862, Ser864, Thr710, Tyr879, Thr880 and Thr992) (Extended Data Fig. 7). When MDIS1-Flag and MIK1-HA were expressed in protoplasts separately, LURE1.2 induced the auto-phosphorylation of MIK1 but not of MDIS1 (Fig. 3g). When MDIS1-Flag and MIK1-HA were co-expressed in the presence of LURE1.2, both MDIS1 and MIK1 were phosphorylated (Fig. 3h). Furthermore, LURE1.2 induces dimerization of MIK1, whereas MDIS1 dimerizes constitutively (Fig. 3i).

The homologues of MDIS1, MIK1 and MIK2 exist in these closely related species. We detected transcripts of CrMDIS1 and EsMDIS1 in the pollen of *C. rubella* and *Eutrema salsagineum*, but not CrMIK1 and EsMIK2 (Extended Data Fig. 8). This suggests that expression of MIK1 and MIK2 in pollen evolved after the divergence between *C. rubella* and the ancestor of *A. thaliana*. This indicates that the MDIS1-MIK complex in the pollen tube was newly evolved, and MDIS1 may function as receptor of the attractants in the older species solely or synergistically with other RLKs. Thus, to explore whether AtMDIS1 is able to break down the reproduction isolation barrier, we transformed AtMDIS1 to *C. rubella*. Using a semi-in-vitro assay, the micropyle targeting efficiency

of transgenic *C. rubella* pollen tubes to the *A. thaliana* ovules is substantially increased (Fig. 4). Since the discovery of LUREs as the female attractant, the search for its male receptor has been hampered by the redundancy of the receptors and LUREs. In this study, we provided strong biochemical, cytological and genetic evidences that the MIK1–MDIS1 complex functions as the LURE1 receptor and determined their activation mechanism. Nevertheless, our data and others also indicate that there are other LURE receptors that are yet to be identified.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 November 2015; accepted 7 January 2016.

Published online 10 February 2016.

- Marton, M. L., Cordts, S., Broadhurst, J. & Dresselhaus, T. Micropylar pollen tube guidance by egg apparatus 1 of maize. *Science* **307**, 573–576 (2005).
- Okuda, S. *et al.* Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* **458**, 357–361 (2009).
- Takeuchi, H. & Higashiyama, T. A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol.* **10**, e1001449 (2012).
- Lorraine, A. E., McCormick, S., Estrada, A., Patel, K. & Qin, P. RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. *Plant Physiol.* **162**, 1092–1109 (2013).
- Qin, Y. *et al.* Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet.* **5**, e1000621 (2009).
- Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. & Gruissem, W. GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* **136**, 2621–2632 (2004).
- Wang, Y. *et al.* Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiol.* **148**, 1201–1211 (2008).
- Muschietti, J., Dircks, L., Vancanneyt, G. & McCormick, S. LAT52 protein is essential for tomato pollen development: pollen expressing antisense LAT52 RNA hydrates and germinates abnormally and cannot achieve fertilization. *Plant J.* **6**, 321–338 (1994).
- Li, H. J. *et al.* POD1 regulates pollen tube guidance in response to micropylar female signaling and acts in early embryo patterning in *Arabidopsis*. *Plant Cell* **23**, 3288–3302 (2011).
- Jones, M. A., Raymond, M. J. & Smirnoff, N. Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in *Arabidopsis*. *Plant J.* **45**, 83–100 (2006).
- Fisher, K. & Turner, S. PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. *Curr. Biol.* **17**, 1061–1066 (2007).
- Hirakawa, Y. *et al.* Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. *Proc. Natl Acad. Sci. USA* **105**, 15208–15213 (2008).
- Sels, J., Mathys, J., De Coninck, B. M., Cammue, B. P. & De Bolle, M. F. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant Physiol. Biochem.* **46**, 941–950 (2008).
- Chang, F., Gu, Y., Ma, H. & Yang, Z. AtPRK2 promotes ROP1 activation via RopGEFs in the control of polarized pollen tube growth. *Mol. Plant* **6**, 1187–1201 (2013).
- Lee, J. S. *et al.* Competitive binding of antagonistic peptides fine-tunes stomatal patterning. *Nature* **522**, 439–443 (2015).
- Li, J. *et al.* BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **110**, 213–222 (2002).
- Santiago, J., Henzler, C. & Hothorn, M. Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. *Science* **341**, 889–892 (2013).
- Sun, Y. *et al.* Structural basis for flg22-induced activation of the *Arabidopsis* FLS2–BAK1 immune complex. *Science* **342**, 624–628 (2013).
- Castells, E. & Casacuberta, J. M. Signalling through kinase-defective domains: the prevalence of atypical receptor-like kinases in plants. *J. Exp. Bot.* **58**, 3503–3511 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Qu for technique assistance in the pollen tube guidance assay. We thank Y. Guo, Quantum Design Inc. China, and core facilities of public technology service centre of Institute of Microbiology and Institute of Genetics and Developmental Biology (Chinese Academy of Sciences) for the MST measurement. We thank J. Zhou and Q. Xie for sharing seeds. This work was supported by the Ministry of Science and Technology of China grants 2013CB945103 to W.-C.Y. and 2015CB910202 to H.-J.L. and the National Natural Science Foundation of China 31330053 and 31221063 to W.-C.Y.

Author Contributions H.-J.L. and W.-C.Y. designed the study, interpreted the results and wrote the paper. T.W. performed most of the experiments. L.L. performed the guidance assay. Y.X. performed the mutants screening. M.-X.Z. performed the LURE1 construction. P.-F.J. performed the cell biology analysis and W.C. performed the qPCR experiments. Y.-C.W. performed mass spectrometry analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.-J.L. (hjli@genetics.ac.cn) or W.-C.Y. (wcyang@genetics.ac.cn).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Plant material. The *Arabidopsis thaliana* wild-type (Col-0), T-DNA insertion mutants *mdis1-1* (*GABI_463E06*), *mdis1-2* (*GABI_090F03*), *mdis2* (*SALK_004879*) and *Capsella rubella* were obtained from ABRC stock centre. *mik1* (*SALK_095005*) and *mik2* (*SALK_061769*) were obtained from J. Zhou. The *E. salsugineum* seeds were obtained from Q. Xie. Plants were grown at 22 °C under long-day conditions (16-h light/8-h dark cycles). For *C. rubella* and *E. salsugineum*, the sterilized seeds were vernalized on the MS media at 4 °C for 30 days and then grown at 22 °C under long-day conditions.

In vitro pollen germination and in vivo tube growth. Pollen tubes were germinated on the germination media (1 mM CaCl₂, 1 mM Ca(NO₃)₂, 1 mM MgSO₄, 0.01% H₃BO₄, 18% sucrose and 0.5% agarose) and cultured for 5 h at 22 °C. The germination ratio was scored under light microscopy. Mean value was calculated from three independent experiments and for each experiment, more than 300 pollen were scored. For *in vivo* tube growth, pollen from the wild-type and mutants were pollinated on the emasculated pistil with mature stigma as reported²⁰. The pistils were collected at 3, 6 and 8 h after pollination and fixed for aniline blue staining. The pollen tubes in the pistil were photographed with Leica M205 microscope. The length of pollen tubes was measured with Image J software (<http://rsb.info.nih.gov/ij/>).

Aniline blue staining and microscopy. Flowers at 12c stage were emasculated and left to grow for 12–24 h to achieve pistil maturation. Then about 20 pollen grains from wild-type or mutant plants, respectively, were dispersed on the stigma papillar cells with a tiny brush. After 24 h, pistils were excised and fixed in Carnoy's fixative (75% ethanol and 25% acetic acid) as reported^{21,22}. The pistils were washed in 50 mM PBS buffer (NaHPO₄/NaH₂PO₄, pH 7.0) three times and immersed in 1 M NaOH overnight for softening. Then after three washes with PBS, the pistil was stained with 0.1% aniline blue (pH 8.0 in 0.1 M K₃PO₄) for 6 h. The stained pistils were observed under Axio Skop2 microscope (Zeiss) equipped with an ultraviolet filter set. Ovules with micropylar guidance defect and the ratio of fertilized ovules to the number of pollen tubes in the style were calculated and the mean values from three independent experiments were compared with that of the wild type.

Generation of constructs and plant transformation. For the dominant-negative constructions, the kinase domains were inactivated by replacing the conserved lysine residue in the intracellular ATP-binding domain with glutamic acid to generate dominant-negative constructs. For the atypical kinase, the intracellular domain was chimaERICally replaced with that of BRASSINOSTEROID INSENSITIVE1 (BRI1)²³ receptor kinase with an inactive kinase domain (K to E substitution). For GFP and GUS reporter expression, genomic sequences containing 2 kb native promoters and the genomic coding sequence for *MDIS1* and *MDIS2* were subcloned into the pCAMBIA1300-GFP binary vector. For complementation of *mik* mutants, full-length coding sequence driven by *LAT52* promoter was cloned into pCAMBIA1300. Similarly, full-length LURE1.2 fused with a C-terminal Flag tag driven by the 35S promoter was cloned into the pCAMBIA1300. For complementation assay, the genomic fused GFP constructs were transformed into the mutant using *Agrobacterium*-mediated floral dip method²⁴. To break down the reproductive isolation barrier, the full-length *MDIS1* coding sequence under the *LAT52* promoter was introduced into *C. rubella* by floral dip method.

Protein purification and pull-down assay. LURE1.2 and PDF2.1 lacking the putative N-terminal signal peptides (71 and 55 amino acids, respectively) were fused N-terminally with a His-tag using pET28a vector (Novagen). Similarly, the ecto-domains of MDIS1, MDIS2, MIK1, MIK2 and PRK3 lacking the predicted signal peptides were fused with an N-terminal GST tag using a pGEX4T-2 vector. The fused proteins were expressed in *Escherichia coli* strain Rosetta DE3 (Stratagene). Cells were grown to an *A*_{600 nm} value of 0.6 at 37 °C and then induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 6 h at 22 °C. The cells were lysed by sonication on ice in lysis buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, Complete Protease Inhibitor Cocktail (Roche) and 1 mg ml⁻¹ lysozyme (Wako). After centrifugation at 12,000 g for 20 min at 4 °C, the supernatants and pellets were collected separately; the pellet was washed three times with the lysis buffer. For LURE1, the insoluble His-LURE1.2 peptides in the inclusion bodies were solved in 1 M urea supplemented with 6 M guanidine-HCl (in Tris-HCl buffer, pH 8.0) for 1 h on ice. Then the peptides were diluted at 1:10 and refolded for 3 days at 4 °C using glutathione (reduced form: oxidized form = 10:1, MERCK) and L-arginine ethyl ester dihydrochloride (Sigma-Aldrich). The folded peptides were dialysed with 3-kDa centrifugal filter (Millipore) and eluted with 50 mM Tris-HCl (pH 8.0) and then used for pull-down, co-IP, protoplasts treatment, pollen tube guidance assays and antibody generation. For purification of GST-tagged ectodo-main of MDIS1, MDIS2, MIK1, MIK2 and PRK3 proteins, cells from 2 l culture

were collected and lysed respectively as described above. The supernatants were used for affinity purification by glutathione agarose beads (GE, 17-0756-01) to avoid extra folding process, although more fused proteins were in the pellets than the supernatant. For GST pull-down assay, the purified proteins were mixed and incubated for 3 h and then subjected to pull-down assay with glutathione agarose beads for 3 h at 4 °C. The beads were collected by centrifugation and then washed five times with buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% Triton X-100 and 0.1% SDS. Finally, the proteins bound on the beads were boiled with 1× SDS sample buffer in 95–100 °C water bath and then subjected to SDS-PAGE and immunoblot with anti-GST (GE Healthcare, 27-4577-01) and anti-His (Santa Cruz) antibody. For mobility shift detection of phosphorylated proteins, phosphatase inhibitor phosphatase (Roche) was added during purification and incubation. Moreover, 50 μM Phos-tag (AAL-107) and 50 μM MnCl₂ was added to the gel according to the manufacturer's procedure. After electrophoresis, the gel was treated with 10 mM EDTA, pH 8.0, for 10 min to remove the Mn²⁺ before immunoblot assay.

Co-IP. Seedlings of *LURE1.2-Flag* transgenic plants were ground to fine powder in liquid nitrogen and solubilized with extraction buffer (0.05 M HEPES-KOH, pH 7.5, 150 mM KCl, 1 mM EDTA, 0.1% Triton X-100 with freshly added proteinase inhibitor cocktail (Roche)). The extracts were centrifuged at 10,000 g for 10 min, and the supernatant was incubated with pre-washed anti-Flag M2 magnetic beads (Sigma-Aldrich, M8823) for 3 h at 4 °C, and then the beads was washed six times with the extraction buffer. The immunoprecipitates were eluted with 3 × Flag peptides. For co-IP in protoplasts, the transformed protoplasts expressing MDIS1-HA, MIK-HA and BAK1-HA were incubated with the purified LURE1.2-Flag or the 200 nM folded His-LURE1.2 purified from *E. coli* for 10 min and lysed for co-IP with pre-washed anti-HA agarose beads (Sigma-Aldrich, A2095). The precipitates were diluted with SDS sample buffer, separated on a 10% SDS-PAGE gel and subjected to immunoblot with the corresponding antibodies (anti-Flag, Sigma-Aldrich, F1804; anti-HA, Santa Cruz, sc-7392; anti-His, Santa Cruz, sc-803). *Arabidopsis* protoplast transformation was performed as reported previously²⁵. For the His-LURE1-protoplast binding assay, the protoplasts incubated with 10 μM LURE1.2 for 5 min, washed three times with the culture buffer and then lysed for SDS-PAGE and immunoblot. For the enhanced interaction between MDIS1 and MIK proteins by LURE1.2, the protoplasts co-transformed with MDIS1-HA and MIK1-Flag were divided into two equal volumes. One was incubated with 0.5 nM LURE1.2 and another with equal volume of 50 mM Tris-HCl (pH 8.0) as mock control for 10 min and subjected to anti-HA immunoprecipitation. For the phosphorylation test, the transformed protoplasts were divided equally into two and incubated for 10 min with 200 nM LURE1.2 or 50 mM Tris-HCl (pH 8.0), respectively. For competition assay, protoplasts expressing MDIS1-HA, MIK1-HA and MIK2-HA were each divided equally into four centrifuge tubes and incubated with purified LURE1.2-Flag. Then active His-LURE1.2 of different concentrations was added to the protoplasts and incubated for 10 min and subsequently co-immunoprecipitated with anti-HA conjugated agarose beads. For co-IP *in planta*, the flowers opened in the morning were collected in the afternoon at the estimated time when the pollen tubes are approaching the ovules. Total proteins were subjected to co-IP with anti-GFP conjugated agarose (ChromoTek, gta-200) or anti-LURE1.2 and protein-A-conjugated magnetic beads (Bio-Rad, 161-4013). The immunoprecipitates was subjected to SDS-PAGE and immunoblot with the corresponding antibodies (anti-GFP-HRP, Miltenyi Biotec, 130-091-833). All the co-IP experiments were repeated at least three times.

Semi-*in-vitro* pollen germination and guidance assay. For *A. thaliana*, the same germination media as that for *in vitro* germination was used. For *C. rubella*, a modified media (4 mM CaCl₂, 4 mM Ca (NO₃)₂, 0.01% H₃BO₄, 10% sucrose and 0.5% agarose) was used. Semi-*in-vitro* germination and ovule-pollen attraction assay were performed as reported in *A. thaliana*³. Pollen tubes entered the micropyle were scored as successful breakdown of the reproductive isolation and the pollen tubes bypass outside the micropyle within 20 μm were scored as failing to enter the micropyle. For the attraction assay, gelatin (Nacalai) beads containing 40 μM LURE1.2 were made and placed beside the pollen tube tip using a micro-manipulator (Narishige) equipped with an inverted microscope (Zeiss AxioVert. A1) as described previously²⁶. Behaviour of pollen tubes was monitored and recorded with a CCD camera. Pollen tubes growing to the beads with >30° direction change were regarded as effective pollen tube attraction.

qPCR. Total RNA was extracted from pollen, *in vitro* germinated pollen tubes (3 h after pollination) and seedlings with TRIzol reagent (Invitrogen) and then treated with DNase I (RNase-free DNase kit, Qiagen) to remove DNA. SuperScript III Reverse Transcriptase (Invitrogen) was used for the reverse transcription reactions. qPCR was performed with Power SYBR Green PCR Master Mix on the Bio-RAD C1000 Thermal Cycler using *Tubulin 2* as the internal control for quantitative

normalization. The specificity of the primers was examined by running the PCR products on 2.5% agarose gels and sequencing.

MST assay. The affinity of the purified GST, GST–MDIS1^{ECD}, MDIS2^{ECD}, MIK1^{ECD}, MIK2^{ECD}, ERECTA^{ECD} and PXY^{ECD} to His–LURE1.2 was measured using the Monolith NT.115 (Nanotemper Technologies). The GST-fusion proteins were fluorescently labelled according to the manufacturer's procedure. The solution buffer was exchanged to labelling buffer and the protein concentration was adjusted to 2 μM. Then fluorescent dye NT-647-NHS was added and mixed and incubated for 30 min at 25 °C in the dark. Finally, the labelled proteins were dialysed with column B (Nanotemper L001) and eluted with 50 mM Tris-HCl (pH 8.0) supplemented with 0.02% Tween 20. For each assay, the labelled protein (about 1 μM) was incubated with the same volume unlabelled His–LURE1.2 of 12 different serial concentrations in 50 mM Tris-HCl (pH 8.0) supplemented with 0.02% Tween 20 at room temperature for 10 min. The samples were then loaded into silica capillaries (Polymicro Technologies) and measured at 25 °C by using 20%–40% LED power and 20% MST power. Each assay was repeated three times. Data analyses were performed using Nanotemper analysis software and OriginPro 9.0 software.

Bimolecular fluorescence complementation analysis in tobacco. The constructs containing MDIS1-NE (MDIS1 fused with the N-terminal YFP), MIK1-CE and MIK2-CE (MIK1 and MIK2 fused with the C-terminal YFP, respectively) were generated as described previously⁸. The *Agrobacterium tumefaciens* EHA105 strains carrying MDIS1-NE and MIK-CE were equally mixed with and without EHA105 strain carrying LURE1.2-Flag and transformed into half of the same tobacco leaf. The transformed leaves were photographed 2 days later with a confocal laser scanning microscope (Zeiss Meta 510). Images were acquired using the same optical setting and average total pixel intensity values were calculated by sampling images of different leaves using the ImageJ software as reported²⁷. Mean values of three experiments, each with five transformed leaves, were compared using Student's *t*-test for biological significance.

Determination of phosphorylation sites and disulphide bonds of MDIS1 and MIK1 *in vitro*. The *E. coli* cells expressing the fusion proteins were lysed and centrifuged at 4 °C. The affinity-purified fusion proteins from the supernatants were subjected to mass spectrometry. His–MDIS1^{KD} was incubated with GST–MIK1^{KD} *in vitro* in kinase assay buffer (25 mM Tris-HCl, pH 8.0, 10 mM MgCl₂ and 100 mM ATP) for 1 h at 30 °C. The proteins were separated by 10% SDS-PAGE and the gel was stained with Coomassie blue G250. The corresponding proteins band were cut into slices and subjected to alkylation/tryptic digestion followed by LC-MS/MS as reported previously²⁸. For disulfide bonds determination, GST–MDIS1^{ECD}, GST–MIK1^{ECD} and GST–MIK2^{ECD} were affinity purified from the supernatants of the bacterial lysis and eluted with 50 mM Tris-HCl, pH 8.0. Then disulfide bonds were determined by mass spectrometry as previously reported²⁹.

Phylogenetic analysis. Alignment of protein sequences were aligned using ClustalW2 program (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Phylogenetic tree of the alignment were drawn with MEGA5 (<http://www.megasoftware.net/>) using the neighbour-joining method with bootstrapping based on 1,000 replicates. The leucine-rich repeat domains were predicted with LRRfinder (<http://www.lrrfinder.com/>) and HHPREP program. The transmembrane domains were predicted with TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). The signal peptides were predicted with SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>).

Yeast two-hybrid assay. The coding sequences of *MDIS1* or *MIK1* and *MIK2*, respectively, were cloned into the pBT3-SUC bait or pPR3-N prey according to the manufacturer's procedure (DualsystemBiotech). Yeast strain NMY51 was co-transformed with the bait and prey constructs and grown on the selective medium lacking Trp, Leu, His and adenine.

RT-PCR. Total RNA was extracted from pollen, leaf, flower and total plant of *C. rubella* and *E. salsugineum* with TRIzol reagent (Invitrogen) and then treated with DNase I (RNase-free DNase kit, Qiagen) to remove any contaminating DNA. SuperScript III Reverse Transcriptase (Invitrogen) was used in reverse transcription

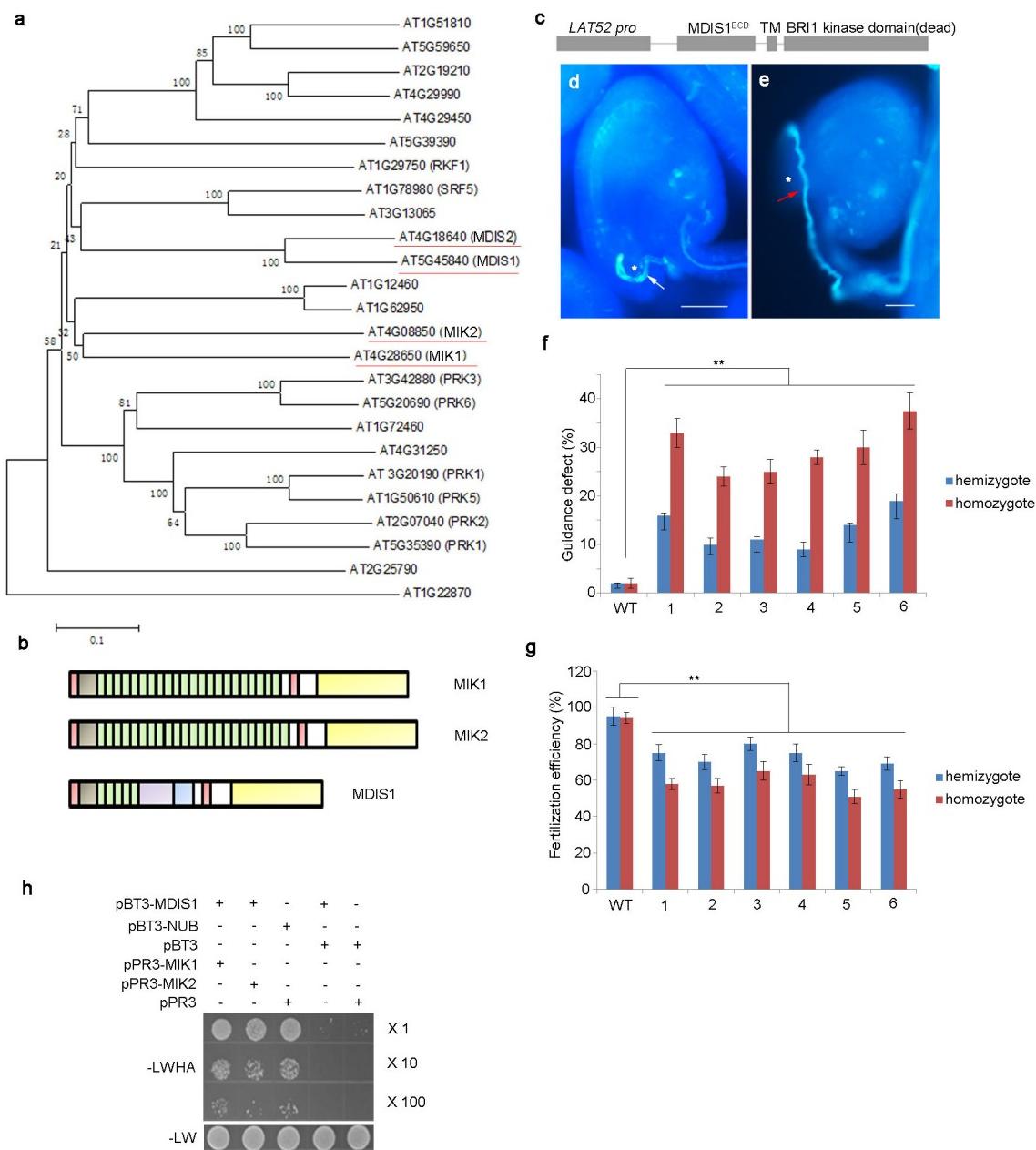
reactions. *ACTIN11* was used as the control for quantitative normalization. The specificity of the primers was confirmed by sequencing of the band after electrophoresis. The accession numbers for the amplified genes are as follows: CrMDIS1 (XM_006280043), EsMDIS1 (XM_006398206), CrMIK1 (XM_006285722), EsMIK1 (XM_006412864), CrMIK2 (XM_006286915), EsMIK2 (XM_006397188), CrACTIN11 (XM_006297859) and EsACTIN11 (XM_006407307).

GUS assay and GFP observation. The histochemical GUS activity assay was performed in the solution containing 2 mM X-Gluc (Sigma) in 50 mM PBS (pH 7.0) and 0.5 mM potassium/ferrocyanide. GUS solution was added to the samples and incubated at 37 °C overnight. Digital images were taken with a Zeiss Axio Skop2 plus microscope. For GFP observation, images were taken with Zeiss confocal laser scanning microscope with a setting of 488 nm excitation (Carl Zeiss, Meta 510 confocal microscope).

Endocytosis of MDIS1–GFP. The semi-*in-vitro* germinated MDIS1–GFP pollen tubes were treated with 500 nM LURE1.2 and photographed by CLSM 780 (Zeiss) after different times.

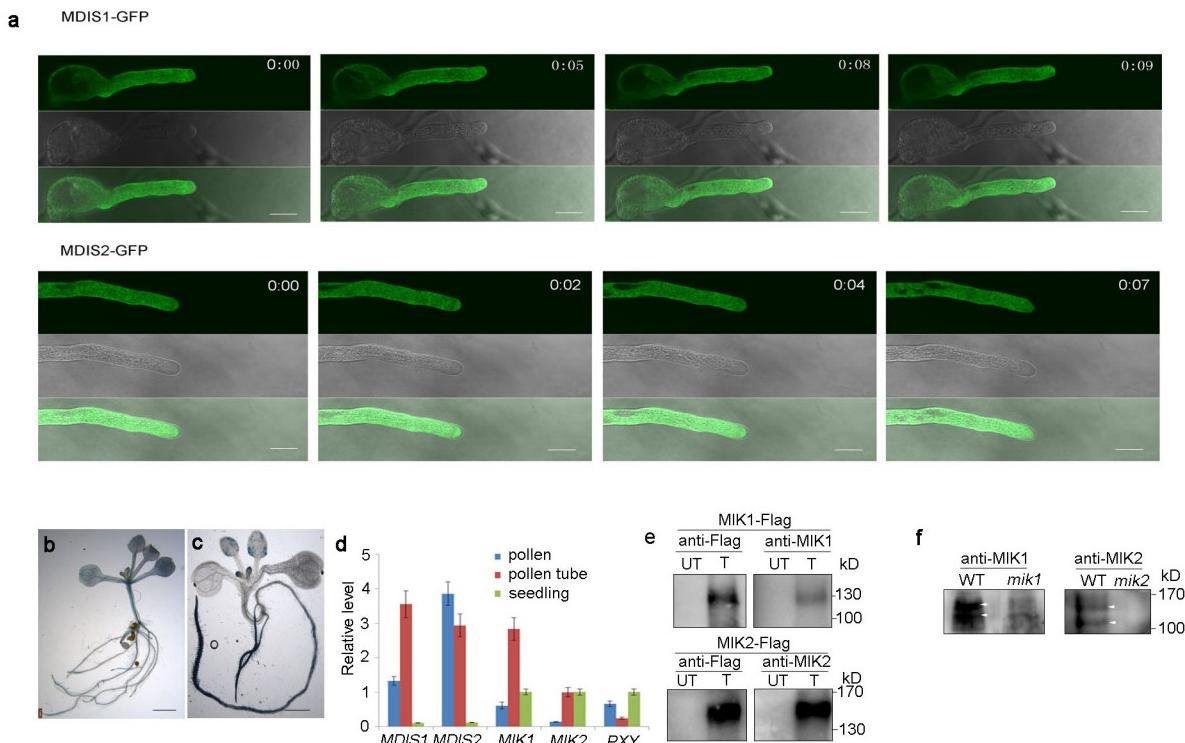
Antibody generation and immunostaining. The anti-MIK1 and anti-MIK2 antibodies were raised in mouse with the purified His-tagged extracellular domains lacking the predicted N-terminal signal peptide. Anti-LURE1.2 antibody was raised in mouse with the folded active His–LURE1.2 fusion protein. For MIK1 and MIK2, the specificity of antibodies was tested with the fusion proteins expressed in protoplasts and the total proteins of pollen from the wild-type and corresponding mutant plants. For LURE1.2, the antibody specificity was tested with the total protein from the leaves of LURE1.2-Flag-overexpressing plants. For immunostaining, the semi-*in-vitro* germinated pollen tubes were fixed in 3.7% paraformaldehyde (3.7% formaldehyde, 1 mM CaCl₂, 1 mM MgSO₄, 50 mM HEPES, 5% sucrose, pH 7.4) for 30 min, washed with PME buffer (50 mM PIPES, 1 mM MgCl₂, 5 mM EGTA, pH 6.8) three times and then subjected to 1% Driselase and 1% cellulase for 10 min. The sample was sequentially washed with PBS buffer (pH 7.4) three times, NP40 buffer (0.5% Nonidet P-40, 1% BSA, in PBS, pH 7.4) and PBS buffer once. Antibodies diluted 1:500 (with PBS containing 3% BSA) were incubated with the sample overnight at 4 °C and then washed with PBS three times. The samples were incubated for 1 h at 4 °C with FITC-labelled goat anti-mouse secondary antibody (KBL, 202-1806) and washed with PBS three times. Anti-fade mounting medium (Invitrogen, P36934) was used for signal detection by confocal laser scanning microscopy (Zeiss Meta 510) with 488 nm excitation.

20. Chen, L. Y. et al. The *Arabidopsis* alkaline ceramidase TOD1 is a key turgor pressure regulator in plant cells. *Nature Commun.* **6**, 6030 (2015).
21. Li, H. J. et al. *Arabidopsis* CBP1 is a novel regulator of transcription initiation in central cell-mediated pollen tube guidance. *Plant Cell* **27**, 2880–2893 (2015).
22. Chen, Y. H. et al. The central cell plays a critical role in pollen tube guidance in *Arabidopsis*. *Plant Cell* **19**, 3563–3577 (2007).
23. Li, J. & Chory, J. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **90**, 929–938 (1997).
24. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
25. Yoo, S. D., Cho, Y. H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* **2**, 1565–1572 (2007).
26. Walter, M. et al. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**, 428–438 (2004).
27. Palikaras, K., Lionaki, E. & Tavernarakis, N. Coordination of mitophagy and mitochondrial biogenesis during ageing in *C. elegans*. *Nature* **521**, 525–528 (2015).
28. Thingholm, T. E., Jorgensen, T. J., Jensen, O. N. & Larsen, M. R. Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nature Protocols* **1**, 1929–1935 (2006).
29. Lu, S. et al. Mapping native disulfide bonds at a proteome scale. *Nature Methods* **12**, 329–331 (2015).



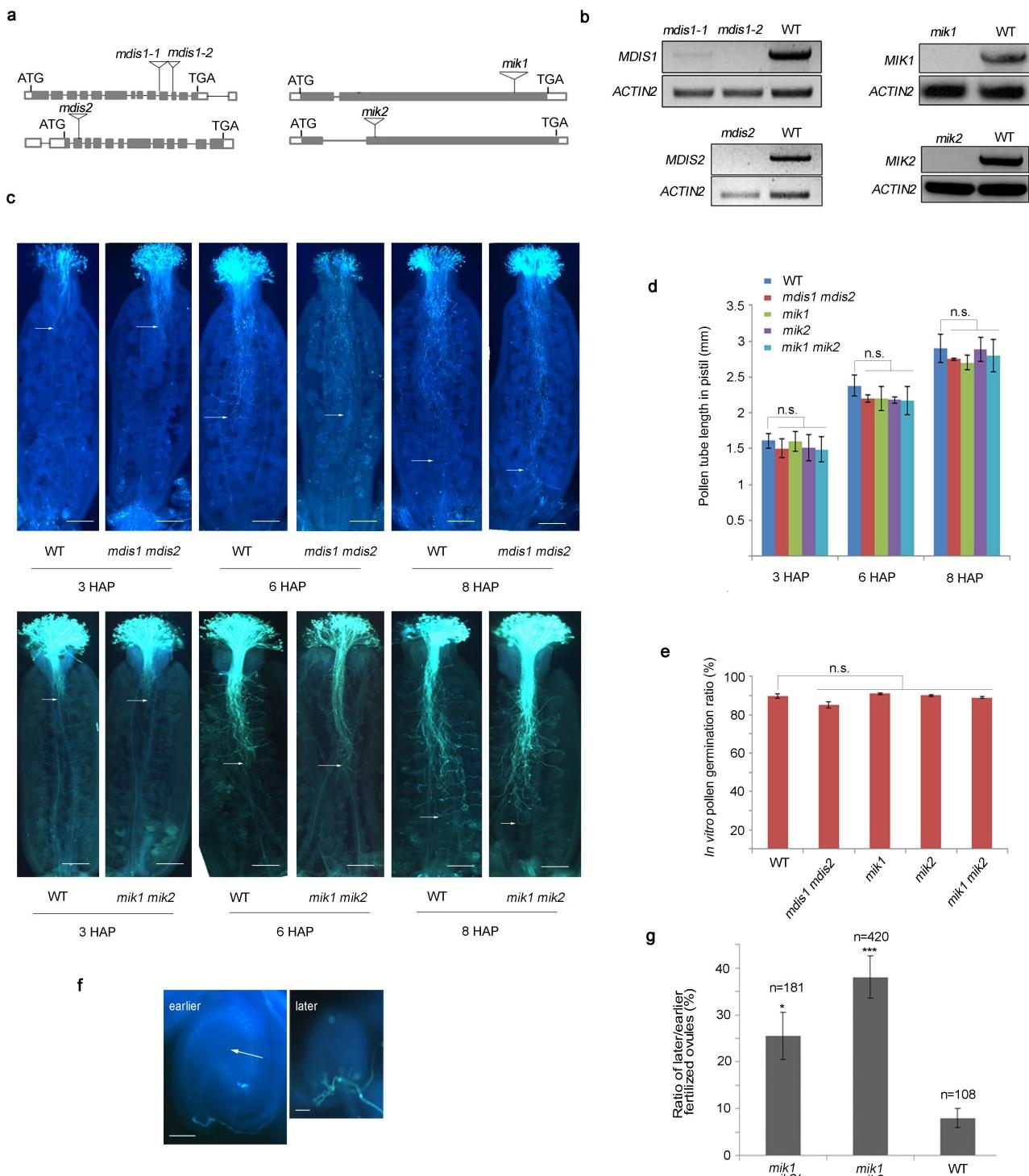
Extended Data Figure 1 | Pollen tubes expressing *MDIS1^{DN}* shows micropylar guidance defect. **a**, Phylogenetic tree of the analysed RLKs expressed in pollen (tubes). **b**, Protein structure of MIK1, MIK2 and MDIS1. Green box, leucine-rich repeats; red, signal peptide and transmembrane domain; yellow, kinase domain; blue, proline-rich domain; purple, linker region. **c**, Schematic diagram of dominant-negative construct of *MDIS1* driven by the pollen-specific promoter *LAT52 pro*. The kinase domain of *MDIS1* was replaced by the dead kinase domain of *BRI1* with an AAG-to-GAG site mutation. **d**, The wild-type pollen tube (arrow) enters the micropyle opening directly. Images are representative of 30 images captured. **e**, The pollen tube (arrow) from the *MDIS1^{DN}* transgenic plants exhibits defective micropylar guidance to the wild-type ovules. Images are representative of 30 images captured. Asterisks in **d** and **e** represent

micropyles. Scale bars, 50 μ m. **f**, Percentage of wild-type ovules with micropylar guidance defect minimally pollinated with pollen from six independent hemizygous and homozygous *MDIS1^{DN}* transgenic lines. Error bars, s.e.m. of 3 independent replicates; **P < 0.01 (Student's t-test); n = 300 for each sample. **g**, Fertilization efficiency of the pollen tubes from the six *MDIS1^{DN}* hemizygous and homozygous lines. The ratio of numbers of successfully targeted pollen tubes to the pollen tubes in the styles was calculated from 30 minimally pollinated pistils. Error bars, s.e.m. of 3 independent replicates; **P < 0.01 (Student's t-test); n = 200 for each sample. **h**, *MDIS1* interacts with MIK1 and MIK2 as shown by dual membrane yeast two-hybrid system. Yeasts were co-transformed with bait construct *MDIS1*-Cub and prey construct *MIK1*-NubG or *MIK2*-NubG, and the transformants were grown on selective media.



Extended Data Figure 2 | MDIS1, MDIS2, MIK1 and MIK2 are expressed in the pollen tubes. **a**, Time-lapse images showing the dynamic distribution of MDIS1-GFP and MDIS2-GFP during pollen tube growth *in vitro*. Images are representative of 30 images captured. Scale bars, 10 µm. **b, c**, Histological GUS staining of seedlings transformed with *MDIS1*- and *MDIS2*-GUS under the native promoters, respectively. Images are representative of 20 images captured. Scale bars, 5 mm. **d**, Quantitative PCR (qPCR) showing the expression of *MDIS1*, *MDIS2*, *MIK1*, *MIK2* and *PXY* in pollen, pollen tubes and seedlings. Error bars, s.e.m of 3 independent replicates. **e**, Specificity test of MIK1 and MIK2

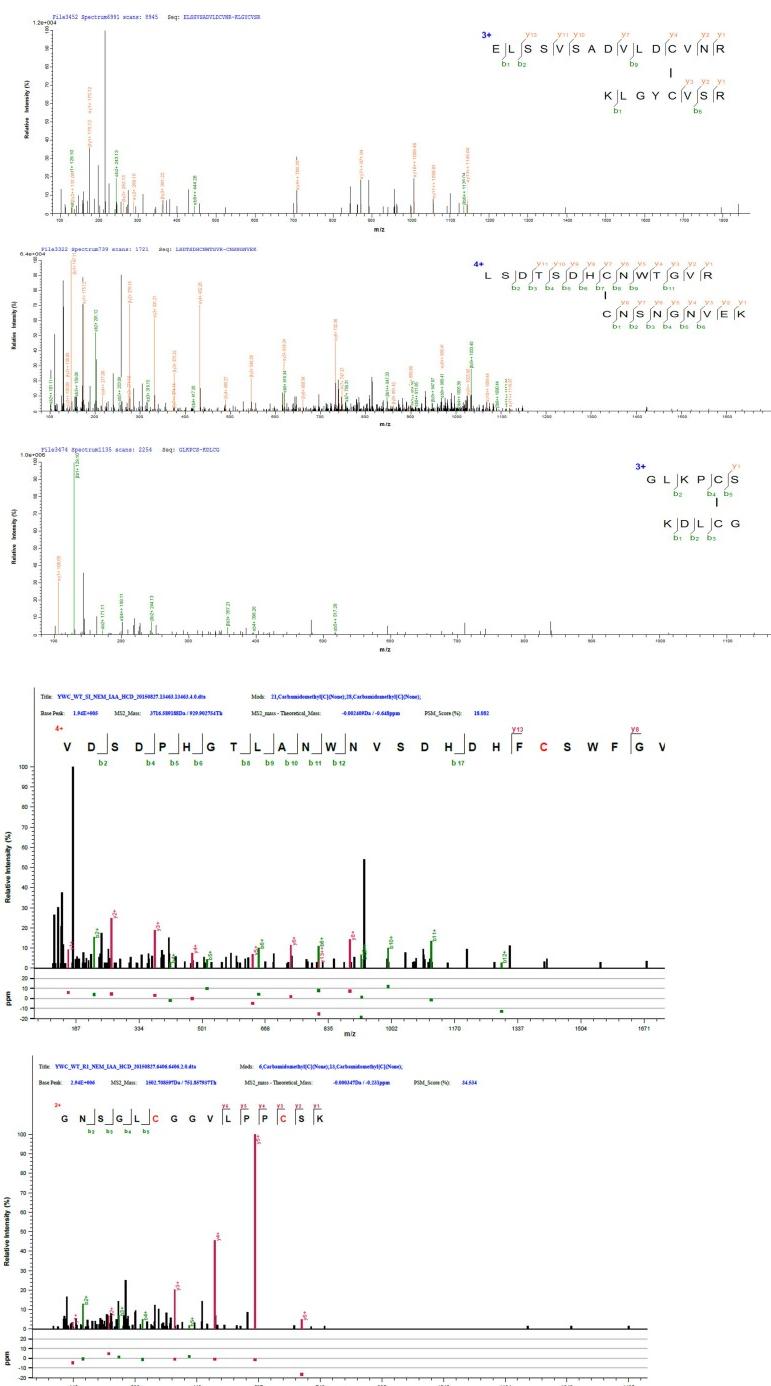
antibodies with *Arabidopsis* protoplasts expressing Flag-tagged MIK1 and MIK2. Equal amount of *Arabidopsis* MIK-Flag-transformed (T) or wild-type (untransformed; UT) protoplasts were lysed and subjected to immunoblotting. Anti-MIK1 and anti-MIK2 recognize the corresponding protoplasts-expressed Flag fusion proteins specifically. **f**, The target protein was recognized by anti-MIK1 and anti-MIK2 in the wild-type pollen, but not in the corresponding mutants. Total protein of the same amount of pollen grains from the wild type and mutants were subjected to SDS-PAGE and immunoblot. Arrows denote target proteins.



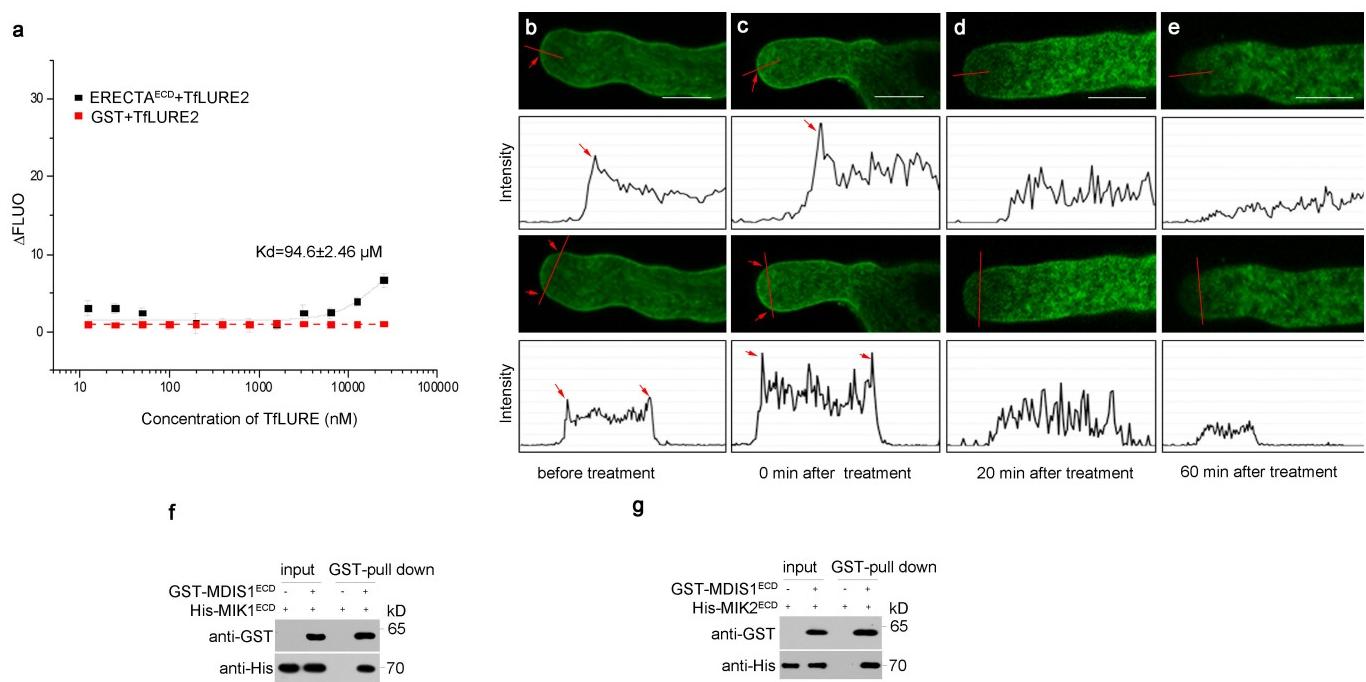
Extended Data Figure 3 | Pollen performance of the mutants.

a, Schematic representation of gene structure of *MDIS1*, *MDIS2*, *MIK1* and *MIK2* and the T-DNA insertion site. The T-DNA insertion positions are indicated by triangles. Filled boxes, exons; open boxes, untranslated region; lines, introns. **b**, Expression of the transcripts in the opening flowers of the wild-type and corresponding mutants. **c**, Representative images of pollen tube length of the corresponding mutants grown in the wild-type pistils at 3, 6 and 8 h after pollination (HAP). Arrows indicate the points the bulk of the pollen tubes reached. Images are representative of 60 images captured. Scale bars, 200 μ m. **d**, Pollen tube length of *mdis1 mdis2*, *mik1*, *mik2* and *mik1 mik2* is comparable to the wild type. $n = 60$ pistils for each sample; $P > 0.1$ (Student's *t*-test); n.s., not significant. Error bars, s.e.m. of 3 independent measurements. **e**, The *in vitro* pollen

germination of *mdis1 mdis2*, *mik1*, *mik2* and *mik1 mik2* is normal. Error bars, s.e.m. of 3 independent replicates; $P > 0.1$ (Student's *t*-test); $n = 300$ for each sample. **f**, **g**, The ratio of earlier to later fertilized wild-type ovules targeted by the *mik1 mik2* and *mik1 mik2/+* pollen tubes is higher than by the wild-type pollen tubes. Approximately 40 pollen tubes were hand-pollinated on the wild-type stigma, which was then subjected to aniline blue staining 30 HAP. Left panel in **f** represents image of the earlier fertilized ovules; right panel represents image of the later fertilized ovules in the same silique. Arrow denotes the enlarged ovule. Scale bars, 20 μ m. **g**, Statistics of results shown in **f**. n , numbers of ovules scored. Error bars, s.e.m. of 3 independent replicates; * $P < 0.05$, *** $P < 0.001$ (Student's *t*-test).

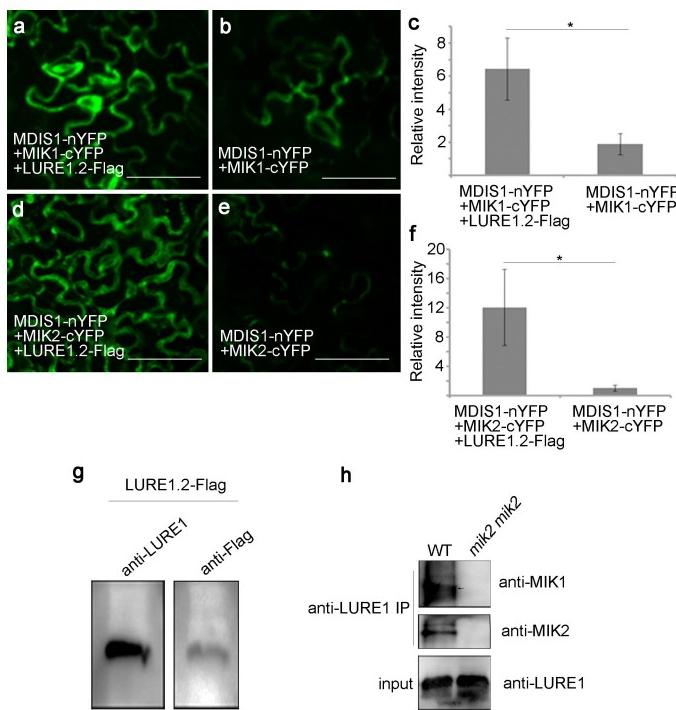


Extended Data Figure 4 | Verification of the predicted disulfide bonds by mass spectrometry. Disulfide bonds of the purified MDIS1^{ECD}, MIK1^{ECD} and MIK2^{ECD} were identified at Cys193–Cys201 of MDIS1, Cys60–Cys67 of MIK1 and Cys683–Cys695 of MIK2. Cys64 of MDIS1, Cys609 and Cys616 of MIK1 were at the oxidized form.

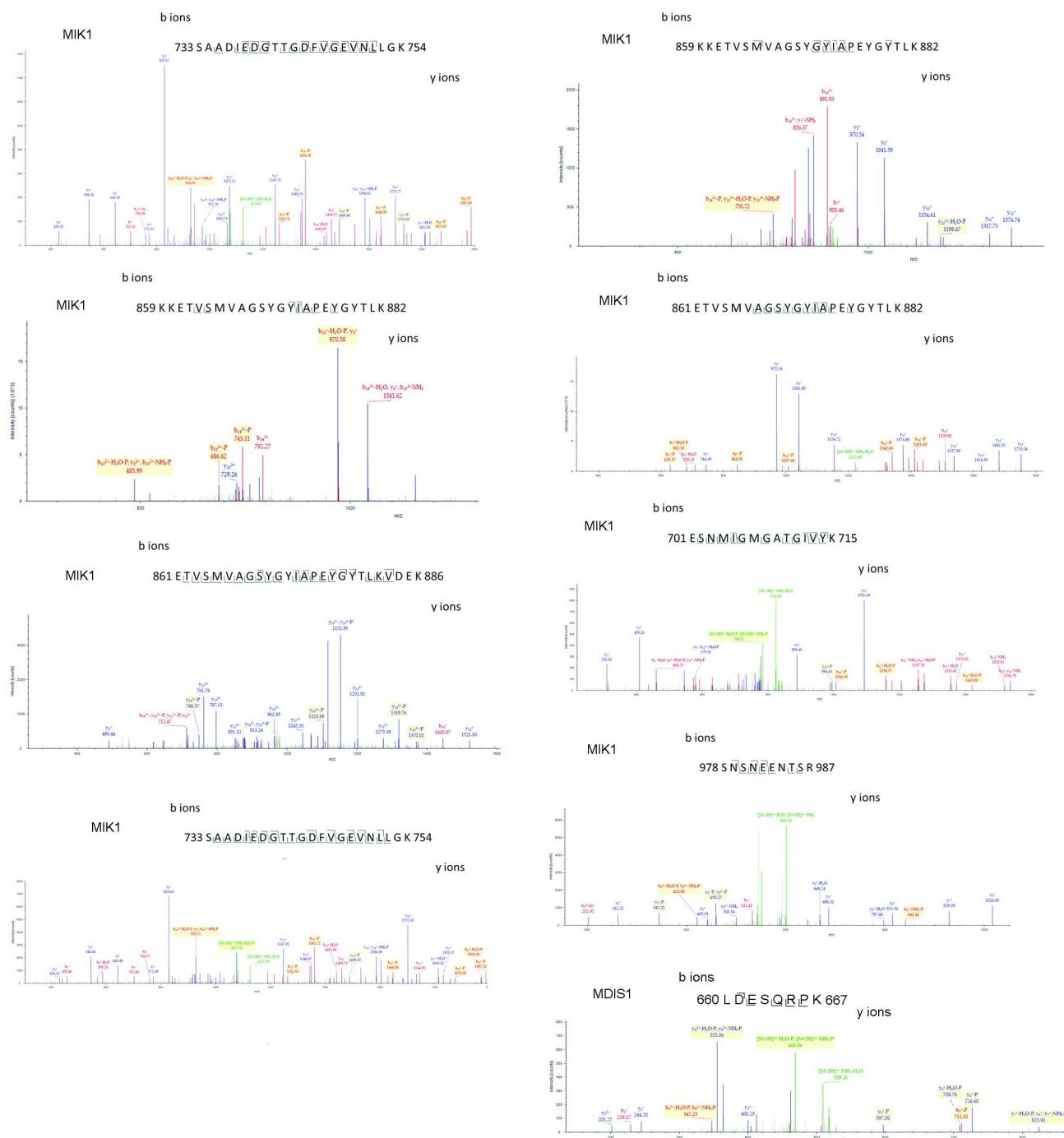


Extended Data Figure 5 | LURE1.2 induced the endocytosis and decrease of MDIS1–GFP in the pollen tube tip. **a**, Binding affinity between ERECTA and TfLURE2 by MST. Error bars, s.e.m. of 3 independent measurements. **b–e**, Confocal images showing the distribution of MDIS1–GFP before LURE1.2 (0.5 μM) treatment (**b**), and at 0 min (**c**), 20 min (**d**) and 60 min (**e**) after treatment. Images are

representative of 63 images captured. Intensity plots along the red lines of each image are shown below. Scale bars, 5 μm. The maximum y-axis values are the same for all intensity plots. The arrows indicate the signal accumulation at the plasma membrane. Scale bars, 5 μm. **f, g**, His–MIK1^{ECD} and His–MIK2^{ECD} specifically bind GST–MDIS1^{ECD}, but not the GST affinity beads. Full blots are shown in Supplementary Data.

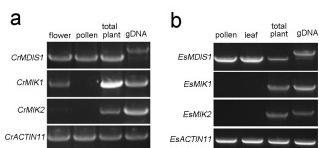


Extended Data Figure 6 | LURE1.2 is perceived by the MDIS1–MIK complex. **a–f.** Confocal images of tobacco leaf showing stronger bimolecular fluorescence complementation signal in the presence of LURE1.2-Flag (**a, d**) as compared with the weak signal in the absence of LURE1.2-Flag (**b, e**). **c–f.** Quantification of the total fluorescence signal of the same areas. Error bars, s.e.m. of 3 independent replicates; * $P < 0.05$ (Student's *t*-test). Five leaves with positive signal were analysed for each experiment. Scale bars, 50 μ m. **g.** Anti-LURE1 and anti-Flag antibodies recognize the LURE1–Flag fusion protein. **h.** Endogenous interaction between LURE and MIK1 or MIK2 by LURE antibody with the total crude proteins extracted from the wild-type pollinated flowers (8 HAP), but not with the *mik1 mik2* mutant. Arrow denotes target proteins. Full blots are shown in Supplementary Data.



Extended Data Figure 7 | Ion-trap MS/MS spectra identifying phosphorylation sites of the kinase domain of MDIS1 and MIK1.
Identification of one phosphorylation site for MDIS1 (Ser663) and eight

for MIK1 (Thr741, Thr742, Thr862, Ser864, Thr710, Tyr879, Thr880 and Thr992) by ion-trap liquid chromatography tandem mass spectrometry (LC-MS/MS).



Extended Data Figure 8 | Expression pattern of homologues of *MDIS1*, *MIK1* and *MIK2* in *C. rubella* and *E. salsugineum* by RT-PCR analysis.

a, *CrMDIS1*, but not *CrMIK1* or *CrMIK2*, is expressed in pollen of *C. rubella*. **b**, *EsMDIS1*, but not *EsMIK1* or *EsMIK2*, is expressed in pollen of *E. salsugineum*. *ACTIN11* transcripts were amplified as controls. Genomic DNA was used as the control for primer specificity.

Extended Data Table 1 | Segregation analysis of *MDIS1^{DN}* represented by segregation ratio of hygromycin resistance (R) to sensitivity (S) with T2 *MDIS1^{DN}* lines carrying a single T-DNA insertion

Parent genotypes		Progeny				
Female	Male	HygR	HygS	R/S	Expected	p
Wild type	<i>MDIS1^{DN}</i> .1/+	97	138	0.7	1	p<0.01
Wild type	<i>MDIS1^{DN}</i> .2/+	111	160	0.69	1	p<0.01
<i>MDIS1^{DN}</i> .1/+	wild type	125	119	1.1	1	NS
<i>MDIS1^{DN}</i> .2/+	wild type	181	190	0.95	1	NS
<i>MDIS1^{DN}</i> .1/+	<i>DIS1^{DN}</i> .1/+	176	76	2.3	3	p<0.01
<i>MDIS1^{DN}</i> .2/+	<i>DIS1^{DN}</i> .2/+	143	64	2.2	3	p<0.01

The *Hpt* gene was introduced to the transformed plants and imports hygromycin resistance when the seedlings were grown on the MS media supplemented with hygromycin. NS, not significant.

Extended Data Table 2 | Transmission efficiency test of *mdis1* and *mdis2* by reciprocal crosses

Parents (♀ X ♂)	Progeny						
	<i>mdis1</i> / <i>MDIS1</i>	<i>MDIS1</i> / <i>MDIS1</i>	<i>mdis2</i> / <i>MDIS2</i>	<i>MDIS2</i> / <i>MDIS2</i>	Total	TE ^F	TE ^M
<i>mdis1</i> / <i>+</i> <i>mdis2</i> / <i>-</i> X WT	100	105	-	-	205	100%	NA
WT X <i>mdis1</i> / <i>+</i> <i>mdis2</i> / <i>-</i>	126	310	-	-	436	NA	40%
<i>mdis1</i> / <i>-</i> <i>mdis2</i> / <i>+</i> X WT	-	-	230	235	465	100%	NA
WT X <i>mdis1</i> / <i>-</i> <i>mdis2</i> / <i>+</i>	-	-	162	192	354	NA	84%

NA, not applicable; TE^F, transmission efficiency of the female gamete; TE^M, transmission efficiency of the male gametes.

Tip-localized receptors control pollen tube growth and LURE sensing in *Arabidopsis*

Hideki Takeuchi^{1,2} & Tetsuya Higashiyama^{1,2,3}

Directional control of tip-growing cells is essential for proper tissue organization and cell-to-cell communication in animals and plants^{1,2}. In the sexual reproduction of flowering plants, the tip growth of the male gametophyte, the pollen tube, is precisely guided by female cues to achieve fertilization³. Several female-secreted peptides have recently been identified as species-specific attractants that directly control the direction of pollen tube growth^{4–6}. However, the method by which pollen tubes precisely and promptly respond to the guidance signal from their own species is unknown. Here we show that tip-localized pollen-specific receptor-like kinase 6 (PRK6) with an extracellular leucine-rich repeat domain is an essential receptor for sensing of the LURE1 attractant peptide in *Arabidopsis thaliana* under semi-*in-vivo* conditions, and is important for ovule targeting in the pistil. PRK6 interacted with pollen-expressed ROPGEFs (Rho of plant guanine nucleotide-exchange factors), which are important for pollen tube growth through activation of the signalling switch Rho GTPase ROP1 (refs 7, 8). PRK6 conferred responsiveness to AtLURE1 in pollen tubes of the related species *Capsella rubella*. Furthermore, our genetic and physiological data suggest that PRK6 signalling through ROPGEFs and sensing of AtLURE1 are achieved in cooperation with the other PRK family receptors, PRK1, PRK3 and PRK8. Notably, the tip-focused PRK6 accumulated asymmetrically towards an external AtLURE1 source before reorientation of pollen tube tip growth. These results demonstrate that PRK6 acts as a key membrane receptor for external AtLURE1 attractants, and recruits the core tip-growth machinery, including ROP signalling proteins. This work provides insights into the orchestration of efficient pollen tube growth and species-specific pollen tube attraction by multiple receptors during male–female communication.

In the final step of pollen tube guidance, two synergid cells on the side of the egg cell are essential for the attraction of the pollen tube to the ovule⁹. We previously identified diffusible and species-specific attractants, defensin-like cysteine-rich LURE peptides, secreted from the synergid cell in the dicot plants *Torenia fournieri* and *A. thaliana*^{4,6}. The attractants of *A. thaliana*, the AtLURE1 peptides, showed considerable attraction activity, but their knockdown partially impaired the precision of the pollen tube guidance around the ovule⁶. Moreover, various additional genes encoding secreted peptides, including many cysteine-rich peptides (CRPs), are likely to be expressed in the female gametophyte¹⁰, suggesting the existence of multiple ligand–receptor pairs for guidance. By focusing on receptor-like kinases (RLKs), which form a large gene family¹¹ and consist of subfamilies with several phylogenetically related genes, we screened T-DNA insertion lines for 23 genes, which encompass almost all pollen-specific RLK genes (see Methods), by a pollen tube attraction assay using the AtLURE1.2 peptide (a representative *A. thaliana* LURE1 peptide)⁶. Under semi-*in-vivo* conditions^{12,13}, pollen tubes from each single mutant grew normally. We found that three independent insertion mutants for PRK6 completely lost their ability to react to AtLURE1.2, whereas all mutants

of the other 22 genes, including seven other PRK genes, reacted to it (Fig. 1a, b). This semi-*in-vivo* result shows that, among the pollen-specific RLKs, PRK6 is essential for pollen tube reorientation towards the AtLURE1 attractant peptide.

PRK6 is one of eight PRK genes¹⁴, which encode transmembrane leucine-rich repeat (LRR) RLKs and are expressed specifically in the pollen tube¹³ (Extended Data Fig. 1a–d). To investigate the subcellular localization of PRK6, we introduced the *pPRK6::PRK6-mRuby2* transgene into the *prk6-1* mutant. Pollen tubes expressing PRK6 tagged to the red fluorescent protein mRuby2 (PRK6-mRuby2) displayed a functional response to AtLURE1 (Fig. 1c, d). In growing pollen tubes, PRK6-mRuby2 was localized predominantly at the plasma membrane of the tip, and detected in cytoplasmic granules with cytoplasmic streaming (Fig. 1e and Supplementary Video 1). These semi-*in-vivo* results show that PRK6 could contribute to the reception of an external AtLURE1 peptide at the pollen tube tip.

Studies of tomato LePRKs have suggested that PRK proteins act as signal-transducing receptors through association/dissociation of two PRK proteins¹⁵, and through interaction with several CRPs secreted from pollen and the pistil for pollen germination and growth stimulation^{16,17}. We thus investigated whether other PRK family proteins function in pollen tube growth and attraction in combination with PRK6. Each *prk* single mutant and most *prk* multiple mutants of various combinations showed near-normal fertility (mean values, 85–100%), whereas triple mutants for PRK3, PRK6 and PRK8, which formed a single subclade (PRK3 subclass; Extended Data Fig. 1a), had a reduced seed set (mean values, 52–74%), and an additional mutation for PRK1, which has a gene structure similar to that of the PRK3 subclass genes, markedly reduced the seed set to ~10% (Extended Data Fig. 1e–g). We then analysed growth and responsiveness to AtLURE1 using a newly developed semi-*in-vivo* assay, in which pollen tubes grown on medium containing AtLURE1.2 peptide showed wavy and swollen tip growth in a concentration- and PRK6-dependent manner (Extended Data Fig. 2a–f and Supplementary Video 2). The AtLURE1-induced wavy morphology indicates a normal physiological response dependent on pollen tube competency¹³, because *in vitro* pollen tubes or semi-*in-vivo* *chx21 chx23* mutant pollen tubes, which show a defect in ovule targeting but not in growth¹⁸, did not respond (Extended Data Fig. 2g, h). In this assay, we revealed that pollen tubes from multiple *prk* mutants that possess both *prk3* and *prk6* mutations exhibited a defect in growth, and more interestingly that *prk1 prk3* double mutations as well as a *prk6* single mutation impaired the response to AtLURE1.2 (Fig. 1f and Extended Data Fig. 2i). Expression of PRK3-mClover in *prk3 prk6*, which showed similar tip localization to that of PRK6, restored the growth defect but not the wavy response to AtLURE1 (Extended Data Fig. 2j, k), suggesting that they regulate different signalling pathways for pollen tube growth rather than redundant signalling pathways.

We investigated *in vivo* PRK functions. Consistent with semi-*in-vivo* results, the *prk3 prk6* double mutant showed slow pollen tube growth

¹Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan. ²JST ERATO Higashiyama Live-Holomics Project, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan. ³Institute of Transformative Bio-Molecules (ITbM), Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan.

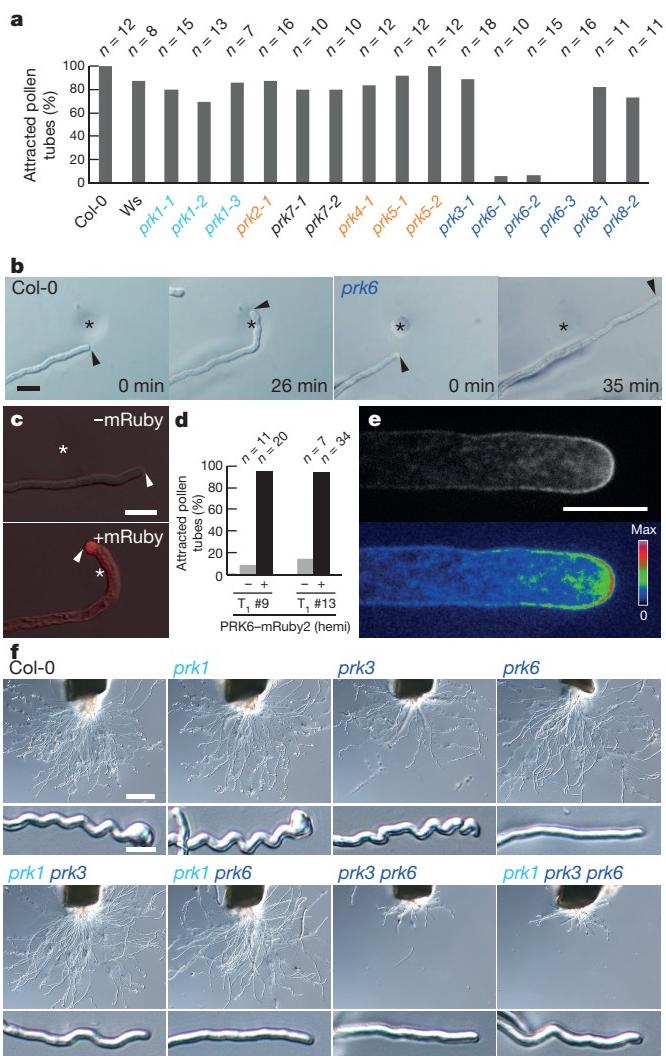


Figure 1 | Pollen tube tip-localized PRK6 and related PRKs are essential for AtLURE1 sensing. **a**, Pollen tube attraction assay for *prk* single mutants using AtLURE1.2. Ws, wild-type *Arabidopsis* ecotype Wassilewskija. **b**, Attracted wild-type (Columbia, Col-0) and insensitive *prk6-1* mutant pollen tubes to AtLURE1 beads (asterisks). Arrowheads mark the tips of the pollen tubes. The data are representative of three images for each of Col-0 and *prk6*; in total, 12 or 10 tubes, respectively, showed similar growth properties. Scale bar, 20 μ m. **c**, **d**, Complementation of the AtLURE1-insensitive *prk6* phenotype by PRK6-mRuby2. In pollen tubes from hemizygous plants, mRuby2-positive (+) but not mRuby2-negative (-) pollen tubes responded to AtLURE1 beads (asterisks). The images are representative of 14 or 3 images for -mRuby or +mRuby, respectively. Scale bar, 20 μ m. **e**, Pollen tube tip localization of PRK6-mRuby2 in a single-plane confocal image (top) and a pseudocolour intensity image (bottom). The data are representative of more than ten tubes. Scale bar, 10 μ m. **f**, Semi-*in-vivo* pollen tube growth/AtLURE1-responsive assay for *prk* mutants 8.5 h after pollination (HAP). The data are representative of at least three assays. Note that, in addition to *prk6*, *prk1 prk3* pollen tubes showed an impaired response to AtLURE1. Scale bars, 100 μ m (top) and 10 μ m (bottom).

as it reached the bottom of the transmitting tract at 24 h after pollination (HAP), compared with 12 HAP in wild-type and *prk6* pollen tubes (Extended Data Fig. 3). We then observed pollen tube attraction towards wild-type ovules on the septum surface. Some pollen tubes of *prk6* single and *prk3 prk6* double mutants, but not the wild type, failed to target nearby ovules (Fig. 2a–c and Extended Data Fig. 4a–d), suggesting that these mutants are less sensitive to ovarian attractants *in vivo*. Furthermore, these mutant pollen tubes showed slightly wandering phenotypes after reaching the ovules, although most ovules

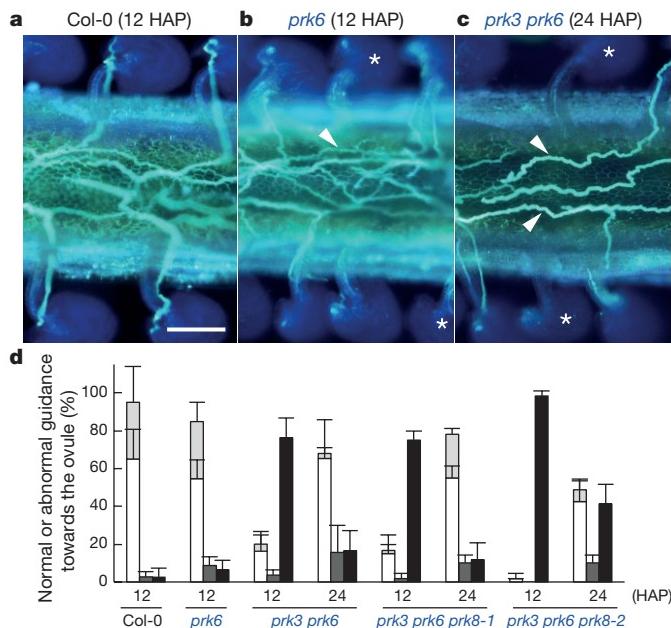


Figure 2 | *prk6* and *prk3 prk6* pollen tubes show decreased ovule-targeting ability. **a–c**, Ovule-targeting of wild-type (**a**), *prk6* (**b**), and *prk3 prk6* (**c**) pollen tubes on the septum surface in wild-type pistils. Asterisks mark ovules that did not attract near pollen tubes passing through them (arrowheads). The data are representative of 1–3 images for each genotype. Similar growth properties were observed in a total of 4 samples. Scale bar, 100 μ m. For entire images of pistils, see Extended Data Fig. 4. **d**, Quantitative analysis of guidance on the ovule. White and light-grey stacked bars show wild-type guidance, in which one pollen tube grows straight on the funiculus and enters the micropyle (white), or an additional pollen tube(s) is associated on the funiculus (light grey). Dark grey bars show abnormal guidance on the ovule, in which a pollen tube takes a 180° turn back on the funiculus and then enters the micropyle. Black bars show ovules that are not associated with a pollen tube. Data are mean and s.d. of four pistils.

eventually attracted mutant pollen tubes (Fig. 2d, dark grey bars), as observed in the AtLURE1-deficient ovules⁶. More severe defects in growth and attraction *in vivo* were observed in *prk3 prk6 prk8* and *prk1 prk3 prk6* triple mutants (Fig. 2d and Extended Data Fig. 4e–g) and were correlated with their fertility. Our physiological analyses demonstrate that the PRK3 subclass and PRK1 could act together as signal-transducing receptors for efficient growth and attraction through sensing of external signalling molecules, including the AtLURE1 attractant peptide.

Next, we examined the intracellular signal transduction mechanism of PRK6. It has been reported that tomato LePRK1 and LePRK2 and *A. thaliana* PRK2 interact with ROPGEF family proteins^{7,8}. ROPGEFs activate intracellular signalling switches, ROPs (Rho-like GTPases from plants), that control various cellular responses^{19–22}. In bimolecular fluorescence complementation (BiFC) assays in tobacco leaf epidermal cells, PRK6 interacted with pollen-expressed ROPGEFs at the plasma membrane (Extended Data Fig. 5a–c). Furthermore, the BiFC assay showed that PRK6 interacted with itself, PRK3, and receptor-like cytoplasmic kinases, LIP1 and LIP2, which are involved in pollen tube growth and attraction and partly in AtLURE1 signalling²³ (Extended Data Fig. 5d). These results indicated that PRK6 forms a complex with factors for proper tip-growth at the plasma membrane.

We then investigated the essential domain of PRK6 for interaction with ROPGEFs and signal transduction using truncated PRK6 proteins (Extended Data Fig. 6a). A co-immunoprecipitation assay demonstrated that a kinase-domain-deleted PRK6 mutant (K-del) as well as full-length PRK6 were associated with ROPGEF12 *in planta*, whereas a cytosolic domain-deleted PRK6 mutant (cyto-del-1) was not (Extended Data Fig. 5e). Corresponding to this, PRK6 (K-del), but

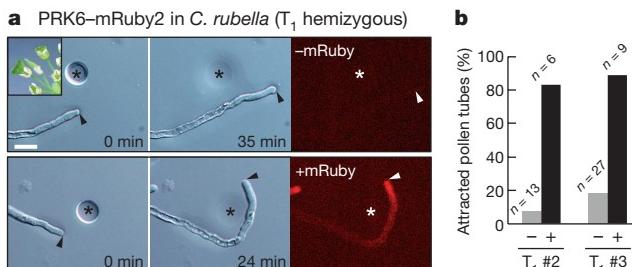


Figure 3 | PRK6 confers the ability to respond to the AtLURE1 peptide on *Capsella* pollen. **a, b,** Attraction assay for *C. rubella* pollen tubes by AtLURE1.2 beads (asterisks). In pollen tubes from T_1 hemizygous *C. rubella* plants, wild-type pollen tubes (top images in **a**; ‘-’ in **b**) did not respond to AtLURE1, whereas pollen tubes containing PRK6-mRuby2 (bottom images in **a**; ‘+’ in **b**) did respond. Arrowheads mark the tips of the pollen tubes. The data are representative of 12 or 9 images for -mRuby or +mRuby, respectively. Scale bar, 20 μm .

not a modified cytosolic domain-deleted PRK6 (cyo-del-2), complemented the AtLURE1-insensitive phenotype of the *prk6* single mutant (Extended Data Fig. 7). Interestingly, PRK6 (K-del) did not restore the growth defect caused by *prk3 prk6* mutations (Extended Data Fig. 7). These results suggest that the membrane-spanning PRK6 interacts with the downstream ROPGEFs via the juxtamembrane domain (the region between the transmembrane and kinase domains) for sensing of AtLURE1, and that the kinase domain of PRK6 has an important role together with PRK3 in pollen tube growth.

Flowering plants have several PRK proteins. Eight orthologous PRK genes were found in the close relatives *Arabidopsis lyrata* and *C. rubella* (Extended Data Fig. 6c). To determine whether PRK6 is sufficient to confer pollen tube responsiveness to a species-specific AtLURE1 peptide, we generated *C. rubella* plants expressing mRuby2-fused *A. thaliana* PRK6, which has a diverged ectodomain compared with the *C. rubella* PRK6 orthologue (CrPRK6) (Extended Data Fig. 6d). In a semi-*in-vivo* assay using *C. rubella* pistils and pollen from T_1 hemizygous *C. rubella* plants, which produce haploid wild-type and transgenic pollen tubes, wild-type *C. rubella* pollen tubes did not react to AtLURE1.2, whereas *C. rubella* pollen tubes expressing *A. thaliana* PRK6 acquired the ability to respond to AtLURE1.2 (Fig. 3a, b). When we used an *A. thaliana* pistil for semi-*in-vivo* growth of *C. rubella* pollen tubes, a similar ability was acquired (4%, $n = 49$ for wild type; 50%, $n = 10$ for PRK6-mRuby2 *C. rubella* in lines #2 and #3). In an opposite manner, we assessed the ability of CrPRK6 to perceive AtLURE1 in *A. thaliana*. Although expression of CrPRK6 restored the AtLURE1-insensitive phenotype of the *prk6* single mutant, it only partially restored the AtLURE1-insensitive phenotype in the *prk* multiple mutants, unlike PRK6 of *A. thaliana* (Extended Data Fig. 7). All of our genetic data indicate that PRK6 acts as a key receptor for sensing of species-specific AtLURE1 attractants in cooperation with other PRKs of *A. thaliana*.

Finally, we tested the hypothesis that tip-focused PRK6 re-localizes to direct tip growth direction towards the AtLURE1 attractant. We performed time-lapse observation of the PRK6-mRuby-expressed pollen tube during reorientation towards Alexa488-labelled AtLURE1.2 (Supplementary Videos 3 and 4). Before and at the time of applying an AtLURE1 bead, PRK6 was observed symmetrically at the tip (Fig. 4a, g). Interestingly, PRK6 accumulated asymmetrically on the AtLURE1 bead side of the tip just before the pollen tube tip growth changed direction (Fig. 4b, c, g). The tip subregion where PRK6 had accumulated expanded gradually to change the growth direction towards AtLURE1 (Fig. 4d–f).

Here, we have shown that pollen tube tip-localized PRK6 regulates the direction of pollen tube tip growth as an essential receptor for AtLURE1 signalling. The pollen tube tip marker PRK6 could be re-localized asymmetrically by the external AtLURE1 peptide and may recruit the intracellular core tip growth machinery, such as ROPGEFs

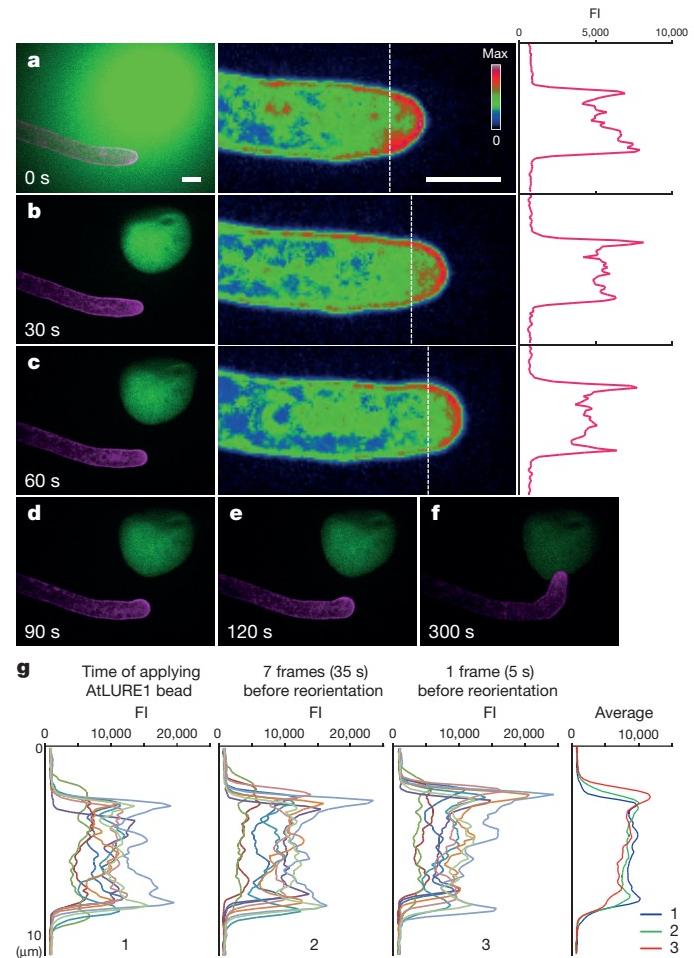


Figure 4 | AtLURE1 induces asymmetric accumulation of tip-localized PRK6 before reorientation of pollen tube tip growth. **a–f,** Time-lapse images of a PRK6-mRuby2 pollen tube during reorientation towards Alexa488-labelled AtLURE1.2. A gelatine bead containing Alexa488-AtLURE1.2 was placed on the medium at 0 s. Higher-magnification fluorescence intensity images of PRK6-mRuby2 are shown in the right panels in **a–c**. Asymmetric accumulation of PRK6-mRuby2 at the tip membrane was observed before a morphological change in the tip towards the AtLURE1 source (**b, c**, right panels). Fluorescence intensity (FI) along a white dashed line (2.5 μm from the front edge) is shown at the right. The sequential data are representative of 10 samples. Scale bars, 5 μm . For full time-lapse images, see Supplementary Videos 3 and 4. **g,** Fluorescence intensity of PRK6-mRuby2 along a line normal to the tip growth axis as shown in the right panels in **a–c**. Data for a further nine pollen tube samples and their mean lines are shown.

and ROP1 (ref. 2), for pollen tube reorientation. Furthermore, our results demonstrate that PRK6, in cooperation with other PRK family receptors, has a central role in the response to species-specific AtLURE1 and mediates efficient pollen tube growth in the pistil. In addition to studies on LePRK2 (refs 15–17), our genetic and physiological data suggest that pollen tube growth and attraction are fine-tuned via interactions among many receptors and multiple stimulants/attractants for successful reproduction. Although PRKs can potentially interact with CRPs^{16,17}, the specific interaction between AtLURE1 and PRKs cannot be established because of AtLURE1 stickiness, which is mediated by a basic amino acid patch of AtLURE1 essential for its activity (Extended Data Fig. 8 and Supplementary Discussion). The sticky property and the attraction activity of AtLURE1 cannot be separated at present. Plants encode many CRPs (>800 genes in *A. thaliana*²⁴) and RLKs with an extracellular domain (>450 genes in *A. thaliana*¹¹). Plant surface receptors have evolved to recognize a variety of CRPs not only for cell differentiation^{25,26},

expansion²⁷ and self-recognition in the pollen–pistil interaction²⁸, but also for positional signals for actively polarizing cells. It will be exciting to explore the molecular basis by which an assembly of receptors determines ligand specificity and to conduct real-time monitoring of ligand-induced signal transduction using the LURE attractant peptide.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 June 2015; accepted 15 February 2016.

1. Itofusa, R. & Kamiguchi, H. Polarizing membrane dynamics and adhesion for growth cone navigation. *Mol. Cell. Neurosci.* **48**, 332–338 (2011).
2. Yang, Z. Cell polarity signaling in *Arabidopsis*. *Annu. Rev. Cell Dev. Biol.* **24**, 551–575 (2008).
3. Higashiyama, T. & Takeuchi, H. The mechanism and key molecules involved in pollen tube guidance. *Annu. Rev. Plant Biol.* **66**, 393–413 (2015).
4. Okuda, S. et al. Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* **458**, 357–361 (2009).
5. Márton, M. L., Fastner, A., Uebler, S. & Dresselhaus, T. Overcoming hybridization barriers by the secretion of the maize pollen tube attractant ZmEA1 from *Arabidopsis* ovules. *Curr. Biol.* **22**, 1194–1198 (2012).
6. Takeuchi, H. & Higashiyama, T. A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol.* **10**, e1001449 (2012).
7. Kaiohni, P. et al. Kinase partner protein interacts with the LePRK1 and LePRK2 receptor kinases and plays a role in polarized pollen tube growth. *Plant J.* **42**, 492–503 (2005).
8. Zhang, Y. & McCormick, S. A distinct mechanism regulating a pollen-specific guanine nucleotide exchange factor for the small GTPase Rop in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **104**, 18830–18835 (2007).
9. Higashiyama, T. et al. Pollen tube attraction by the synergid cell. *Science* **293**, 1480–1483 (2001).
10. Jones-Rhoades, M. W., Borevitz, J. O. & Preuss, D. Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. *PLoS Genet.* **3**, e171 (2007).
11. Shiu, S. H. & Bleecker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. USA* **98**, 10763–10768 (2001).
12. Palanivelu, R. & Preuss, D. Distinct short-range ovule signals attract or repel *Arabidopsis thaliana* pollen tubes *in vitro*. *BMC Plant Biol.* **6**, 7 (2006).
13. Qin, Y. et al. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet.* **5**, e1000621 (2009).
14. Chang, F., Gu, Y., Ma, H. & Yang, Z. AtPRK2 promotes ROP1 activation via RopGEFs in the control of polarized pollen tube growth. *Mol. Plant* **6**, 1187–1201 (2013).
15. Wengier, D. et al. The receptor kinases LePRK1 and LePRK2 associate in pollen and when expressed in yeast, but dissociate in the presence of style extract. *Proc. Natl Acad. Sci. USA* **100**, 6860–6865 (2003).
16. Tang, W., Ezcurra, I., Muschietti, J. & McCormick, S. A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2. *Plant Cell* **14**, 2277–2287 (2002).
17. Tang, W., Kelley, D., Ezcurra, I., Cotter, R. & McCormick, S. LeSTIG1, an extracellular binding partner for the pollen receptor kinases LePRK1 and LePRK2, promotes pollen tube growth *in vitro*. *Plant J.* **39**, 343–353 (2004).
18. Lu, Y. et al. Pollen tubes lacking a pair of K⁺ transporters fail to target ovules in *Arabidopsis*. *Plant Cell* **23**, 81–93 (2011).
19. Berken, A., Thomas, C. & Wittinghofer, A. A new family of RhoGEFs activates the Rop molecular switch in plants. *Nature* **436**, 1176–1180 (2005).
20. Oda, Y. & Fukuda, H. Initiation of cell wall pattern by a Rho- and microtubule-driven symmetry breaking. *Science* **337**, 1333–1336 (2012).
21. Li, H., Lin, Y., Heath, R. M., Zhu, M. X. & Yang, Z. Control of pollen tube tip growth by a Rop GTPase-dependent pathway that leads to tip-localized calcium influx. *Plant Cell* **11**, 1731–1742 (1999).
22. Gu, Y. et al. A Rho family GTPase controls actin dynamics and tip growth via two counteracting downstream pathways in pollen tubes. *J. Cell Biol.* **169**, 127–138 (2005).
23. Liu, J. et al. Membrane-bound RLCKs LIP1 and LIP2 are essential male factors controlling male-female attraction in *Arabidopsis*. *Curr. Biol.* **23**, 993–998 (2013).
24. Silverstein, K. A. et al. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.* **51**, 262–280 (2007).
25. Lee, J. S. et al. Direct interaction of ligand-receptor pairs specifying stomatal patterning. *Genes Dev.* **26**, 126–136 (2012).
26. Lee, J. S. et al. Competitive binding of antagonistic peptides fine-tunes stomatal patterning. *Nature* **522**, 439–443 (2015).
27. Haruta, M., Sabat, G., Stecker, K., Minkoff, B. B. & Sussman, M. R. A peptide hormone and its receptor protein kinase regulate plant cell expansion. *Science* **343**, 408–411 (2014).
28. Takayama, S. et al. Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. *Nature* **413**, 534–538 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Hasebe and S. Miyazaki for seeds of a part of mutant lines; Y. Matsubayashi, H. Shinohara, D. Maruyama, M. Ohtsu and K. Motomura for valuable comments and discussions; D. Kurihara, Y. Hamamura and S. Nagahara for technical assistance with confocal microscopy and physiological analyses; M. M. Kanaoka for agro-infiltration method using tobacco leaf; S. Oishi for reverse-phase high-pressure liquid chromatography (HPLC); and the Japan Advanced Plant Science Network for use of some microscopes. This work was supported by grants from the Japan Science and Technology Agency (ERATO project to T.H.) and the Japan Society for the Promotion of Science Fellowships (no. 5834 to H.T.).

Author Contributions H.T. designed the study; H.T. performed experiments; H.T. and T.H. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.H. (higashi@bio.nagoya-u.ac.jp) or H.T. (hidenori.takeuchi@gmi.oewa.ac.at).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Plant materials. *Arabidopsis thaliana* accession Columbia (Col-0) was used as the wild type. Seeds of T-DNA insertion lines were obtained from ABRC and NASC, and T-DNA insertions were confirmed by genomic PCR (Extended Data Table 1). The insert sites were determined by sequencing of the PCR products, as described in Extended Data Fig. 1c. Plant growth conditions and transformation methods were described previously⁶. *C. rubella* seeds were obtained from ABRC (accession CS22697; ref. 29), and *C. rubella* plants in the rosette stage were subjected to vernalising cold treatment (8-h photoperiod at 4 °C for about 1 month) for flowering induction.

Collection of T-DNA insertion mutants of pollen-expressed RLKs. To investigate candidate RLKs responsible for AtLURE1 signalling, RLK genes encoding proteins with extracellular domains and displaying notable and specific expression in the pollen tube were selected as follows. Whether the more than 80 genes expressed in dry pollen or pollen tubes¹³ were expressed predominantly in the mature pollen was determined using the *Arabidopsis* eFP Browser (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>)³⁰. Twenty-three pollen-dominant genes and their related genes were selected: PRK1-8 (see Extended Data Table 1), AT2G18470 (PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASE 4, PERK4), AT4G34440 (PERK5), AT3G18810 (PERK6), At1g49270 (PERK7), AT1G10620 (PERK11), AT1G23540 (PERK12), AT4G29450, AT3G13065 (STRUBBELIG-RECEPTOR FAMILY 4, SRF4), AT1G78980 (SRF5), AT4G18640 (MORPHOGENESIS OF ROOT HAIR 1, MRH1), AT5G45840, AT1G29750 (RECEPTOR-LIKE KINASE IN FLOWERS 1, RKF1), AT3G23750 (BAK1-ASSOCIATING RECEPTOR-LIKE KINASE 1, BAK1; or TMK4), AT1G19090 (CYSTEINE-RICH RLK 1, CRK1) and AT4G28670. A further five RLK genes of a subclass of the CrRLK1L family (AT3G04690 (ANXUR1), AT5G28680 (ANXUR2), AT4G39110, AT2G21480 and AT5G61350) were also pollen-dominant but were not examined in this study. T-DNA insertions in the coding or promoter regions of these selected 23 genes were identified by genomic PCR and sequencing of the PCR products. Semi-*in-vivo* pollen tubes from one or more lines for each gene were assessed by an attraction assay using the AtLURE1.2 peptide, as described below.

Semi-*in-vivo* attraction assay. Recombinant His-tagged AtLURE1.2 peptide was expressed in *Escherichia coli*, purified and refolded, as described previously⁶. The refolded His-AtLURE1.2 peptide was suggested to be a conformational isomer by reverse-phase high-pressure liquid chromatography (HPLC) using a Phenomenex Jupiter C18 column and a Jasco analytical instrument equipped with a UV-2077 plus detector and PU-2080 plus pumps. A construct for His-AtLURE1.2(GGGG) was generated from pET-28a-AtLURE1.2 by site-directed mutagenesis using the primers 5'-GTATGgGAgGGGGTggGTATATTCT-3' and 5'-CACCCcTCcCAT ACAAGCTC-3' (lowercase bases denote mutated bases from the original AtLURE1.2). No aggregation due to inappropriate folding was observed during refolding or concentration of the His-AtLURE1.2(GGGG) peptide. Alexa488-labelled His-AtLURE1.2 was produced using the refolded His-AtLURE1.2 peptide and the Alexa Fluor 488 Protein Labelling Kit (Thermo Fisher Scientific), according to the manufacturer's protocol. For the semi-*in-vivo* attraction assay, pollen tubes were grown through cut styles of *A. thaliana* on solid pollen germination medium poured into a mould made with 2-mm thick silicone rubber and cover glasses³¹. About 4–5 h after hand-pollination, the topside cover glass was removed and the medium was covered with hydrated silicone oil (KF-96-100CS; Shin-Etsu). The assay for T₁ hemizygous *C. rubella* plants was performed similarly using *A. thaliana* or *C. rubella* pistils as pollen acceptors. Attraction of pollen tubes towards the peptide was evaluated using gelatine beads (5% (w/v) gelatine (Nacalai) in the pollen medium without agar) containing 5 μM His-tagged AtLURE1.2 peptide under an inverted microscope (IX71, Olympus) equipped with a micro-manipulator (Narishige), as described previously⁶. The percentages of attracted pollen tubes are shown for the total number of pollen tubes in at least two assays. In the assay using hemizygous plants, the presence of the transgene in the pollen tube containing the transgene was confirmed by fluorescence observations after assessment of pollen tube responsiveness as a simple blind test. For the AtLURE1-responsive wavy assay, the purified AtLURE1.2 peptide was added to solid pollen germination medium, which was melted at 70 °C and then cooled to a certain degree. The mixture was mixed by vortexing and poured into the mould. Pollen tubes of each genotype were grown through cut styles, as described earlier.

Binary vector construction, genetic transformation, and selection of transformants. Plasmids encoding green and red fluorescent proteins, pcDNA3-Clover and pcDNA3-mRuby2 (gifts from M. Lin, Addgene plasmids 40259 and 40260)³², respectively, were used as templates to prepare binary vectors as follows. The original Clover was converted to A206K mutant form to prevent potential

dimerization, and a restriction site KpnI in the nucleotide sequence was eliminated by a silent mutation, designated as monomeric Clover (mClover). Modified binary vectors pPZP211, pPZP221 (ref. 33) and pMDC99 (ref. 34) derivatives, pPZP211G (ref. 35), pPZP221G, and pMDC99G, were used for cloning of the mClover and mRuby2. pPZP221G was produced by the same procedure as that used for pPZP211G (ref. 35), and pMDC99G was produced by removal of ccdB by EcoRI digestion and self-ligation³¹ and by inserting multiple cloning sites, green fluorescent protein (GFP), and the NosT cassette of pPZP211G via HindIII and EcoRI sites. To add linkers to both the amino-terminal and carboxy-terminal of mClover and mRuby2, three rounds of PCR were performed with DNA templates for mClover and mRuby2, respectively, using three sets of primers: (5'-aggttggatggaaATGGTGAGCAAGGGCGA-3' and 5'-tcacccatccacctaCTTGATACAGCTCGTCCA-3'; 5'-tctggagggtggagggttcAGGTGGAGGTGGA-3' and 5'-cggtttacccacttgttaataagaattcTCCACCTCCACCTG-3'; 5'-aggccgcgcTCTGGAGGTGAG-3' and 5'-cggtttacccacttgttaataagaattcTCCACCTCCACCTG-3') (lowercase bases denote additional nucleotides for template DNAs). The PCR fragments were digested with AscI and KpnI and ligated into pPZP211G, pPZP221G and pMDC99G by replacing the GFP sequence, resulting in pPZP211Clo, pPZP221Clo, pPZP211Ru, pPZP221Ru, pMDC99Clo and pMDC99Ru vectors.

For the expression of full-length PRK6, kinase domain-deleted PRK6 (K-del), cytosolic domain-deleted PRK6 (cyo-del-2) and PRK6 orthologue of *C. rubella* (CrPRK6) as mRuby2-fusion protein under the control of their own promoter, genomic sequences of PRK6 or CrPRK6 containing promoter and coding regions were amplified and were cloned into the pPZP221Ru using SalII and AscI sites, resulting in pPZP221-pPRK6::PRK6-mRuby2, -pPRK6::PRK6 (K-del)-mRuby2, -pPRK6::PRK6 (cyo-del-2)-mRuby2, and -pCrPRK6::CrPRK6-mRuby2 vectors. These constructs were introduced into prk6-1, prk3-1 prk6-1, prk3-1 prk6-1 prk8-2 and prk1-2 prk3-1 prk6-1 plants by the floral dip method. For the heterologous expression of PRK6 in *C. rubella*, the pPZP221-pPRK6::PRK6-mRuby2 vector was used for *C. rubella* transformation by the floral dip method after flowering induction. Genomic sequences of PRK6 or PRK3 containing promoter and coding regions were also cloned into pMDC99Clo using SalII and AscI sites, and these constructs were introduced into prk3-1 prk6-1. Primers used for these constructs are listed in Supplementary Table 1.

For all transgenic lines expressing PRK proteins, T₁ transformants were screened by moderate or weak fluorescence intensity in approximately half of the pollen grains, implying single insertion. Note, when pollen grains showing mid to strong fluorescence intensity were used for the semi-*in-vivo* pollen tube growth assay, few or no fluorescent pollen tubes emerged from the cut end, probably owing to the growth defect caused by excess PRK expression. T₂ homozygous plants obtained from several selected T₁ lines were used for the semi-*in-vivo* AtLURE1-responsive wavy assay.

BiFC assay. To prepare constructs for the BiFC assay in the leaf epidermal cells of *Nicotiana benthamiana*, cauliflower mosaic virus 35S promoter was introduced to the binary vector pPZP211G (ref. 35) using HindIII and PstI sites. Then, the GFP sequence was replaced by nucleotide sequences encoding each of amino acids 1–174 and 175–239 of enhanced yellow fluorescent protein (nYFP and cYFP, respectively) with the same linkers as the mClover and mRuby2 constructs, described above, resulting in pPZP211-p35SnY and pPZP211-p35ScY vectors. Genomic PRK2 and PRK6 were amplified and connected upstream of the cYFP sequence of pPZP211-p35ScY. The genomic sequences of PRK6, PRK3, LIP1 and LIP2 were connected upstream of the nYFP sequence of pPZP211-p35SnY. Genomic ROPGEF8, ROPGEF9, ROPGEF12, ROPGEF13 and ROPGEF12ΔC (encoding amino acids 1–443 of ROPGEF12 (ref. 8)) were amplified and connected downstream of the nYFP sequence in pPZP211-p35SnY. Primers used for these constructs are listed in Supplementary Table 1.

Transient expression in *N. benthamiana* leaves was performed by agro-infiltration according to a method described previously²⁰. In brief, *Agrobacterium tumefaciens* strains GV3101 (pMP90) containing each expression vector were cultured overnight in LB media. Equal amounts of *Agrobacterium* cultures for nYFP and cYFP constructs and the p19 silencing suppressor were mixed to a final $A_{600\text{nm}}$ of 1.0 and collected and resuspended in infiltration buffer (10 mM MES, pH 5.6, 10 mM MgCl₂ and 150 μM acetosyringone). The mixed suspensions were incubated at room temperature for ~3 h and infiltrated into leaves of *N. benthamiana* grown at 25 °C. Two to three days after infiltration, the leaves were cut into pieces for confocal microscope observation.

Analyses of pollen tube growth and guidance in pistils. To analyse pollen tube growth and guidance in the pistil, Col-0 pistils emasculated 1 day before were abundantly hand-pollinated with two or three fully dehiscent anthers from each genotype. Two types of aniline blue staining were performed 12 or 24 h after pollination as follows. For measurement of pollen tube growth inside the transmitting tract, aniline blue staining was performed, as described previously³⁶. Pollinated

pistils were dissected to remove a pair of ovary walls and then fixed in a 9:1 mixture of ethanol and acetic acid for more than 2 h. They were washed with 70% ethanol for ~30 min, treated with 1 N NaOH overnight, and stained with aniline blue solution (0.1% (w/v) aniline blue, 0.1 M K₃PO₄) for more than several hours. The pistils were observed under ultraviolet illumination using an upright microscope (DP71, Olympus). Multiple images for each pistil were combined using Adobe Photoshop CS4 (Adobe Systems), and lengths from the top of the stigma to the tip of the longest pollen tube were measured for maximum pollen tube length using the MacBiophotonics ImageJ software (<http://www.macbiophotonics.ca/>).

To evaluate pollen tube guidance after emergence on the septum surface of the pistil, dissected pistils were stained directly with modified aniline blue solution (5:8:7 (v/v) mixture of 2% aniline blue, 1 M glycerol, pH 9.5, and water), as described previously³⁷, and observed under ultraviolet illumination using an upright microscope (DP71, Olympus). Quantitative analysis was performed by evaluating pollen tube growth on 10 upper ovules of both sides (total, 20 ovules per pistil) to eliminate bias in ovule number in a pistil.

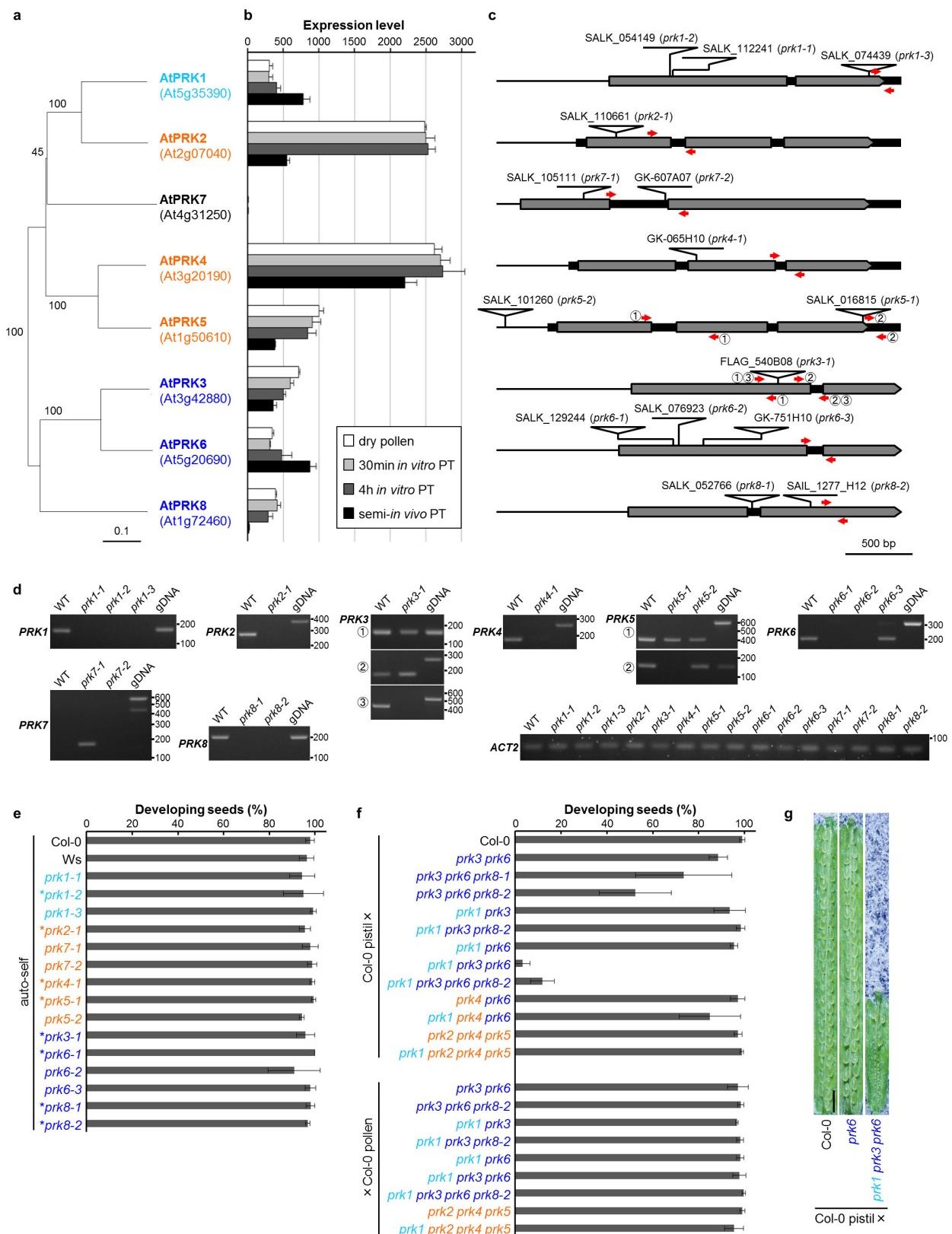
Confocal microscopy. Confocal images were acquired using an inverted microscope (IX81, Olympus) equipped with a spinning disk confocal scanner (CSU-X1, Yokogawa Electric Corporation), 488 nm and 561 nm LD lasers (Sapphire, Coherent), and an EM-CCD camera (Evolve 512, Photometrics). For *A. thaliana* pollen tubes, a 60× silicone immersion objective lens (UPLSAPO60XS, Olympus) and a 1.6× intermediate magnification changer were used. For time-lapse imaging of PRK6-mRuby2 during pollen tube attraction towards a gelatine bead containing 5 μM Alexa488-labelled His-AtLURE1.2, sequential images using 488 nm and 561 nm lasers were acquired every 5 s. For the BiFC assay in *N. benthamiana* leaves, a 20× objective lens (UPLFLN20X, Olympus) was used. The confocal microscope system was controlled and time-lapse images were processed by MetaMorph (Universal Imaging). Images were edited with MacBiophotonics ImageJ.

Co-immunoprecipitation assay. To prepare transient expression vectors in *N. benthamiana* leaf cells, the cauliflower mosaic virus 35S promoter was introduced into the binary vector pPZP211Clo via HindIII and PstI sites, resulting in the pPZP211-p35SClo vector. The 3×Flag tag sequence was introduced into pPZP211-p35S using the AscI and SacI sites, resulting in the pPZP211-p35SFlag vector.

For co-immunoprecipitation of PRK-mClover and ROPGEF12-3×Flag proteins, genomic sequences of full-length PRK3, full-length PRK6, PRK6 (K-del), PRK6 (cyto-del-1), and ROPGEF12 were inserted into the pPZP211-p35SClo or pPZP211-p35SFlag vectors. One of the PRK-mClover or mClover proteins plus the p19 silencing suppressor and ROPGEF12-3×Flag were co-expressed in *N. benthamiana* leaves as described for the BiFC assay. The leaves were ground in mortars with liquid nitrogen and suspended in 3–3.5×(w/v) extraction buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10% glycerol, protease inhibitor cocktail (cComplete EDTA-free, Roche)). The extracts were centrifuged twice at 10,000g for 10 min at

4°C to remove precipitates. The supernatants, with the exception of the mClover sample, were ultracentrifuged at 100,000g for 30 min at 4°C, and the pellets were solubilized in extraction buffer containing 0.5% Triton X-100. The solubilised membrane fraction samples and mClover sample plus 0.5% Triton X-100 were incubated with GFP-trap agarose beads (ChromoTek, gta-20) with rotation for 2 h at 4°C. The beads were washed with buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl) four times. Then, the bound proteins were eluted with SDS sample buffer by heating at 70°C for 5 min. The protein samples were separated on SDS-PAGE and subjected to immunoblot analysis. The immunoblot analysis was conducted on PVDF membranes (Immobilon-P, Millipore) using primary antibodies (anti-GFP (ab290, Abcam), or monoclonal anti-DYKDDDDK tag (Wako) for Flag tag) and secondary antibodies (goat anti-rabbit IgG peroxidase-labelled antibody or goat anti-mouse IgG peroxidase-labelled antibody (KPL)). Signals were visualized using Immobilon Western Chemiluminescent HRP Substrate (Millipore), detected with Light-Capture (ATTO).

29. Slotte, T. et al. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genet.* **45**, 831–835 (2013).
30. Winter, D. et al. An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* **2**, e718 (2007).
31. Hamamura, Y. et al. Live imaging of calcium spikes during double fertilization in *Arabidopsis*. *Nature Commun.* **5**, 4722 (2014).
32. Lam, A. J. et al. Improving FRET dynamic range with bright green and red fluorescent proteins. *Nature Methods* **9**, 1005–1012 (2012).
33. Hajdukiewicz, P., Svab, Z. & Maliga, P. The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol. Biol.* **25**, 989–994 (1994).
34. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469 (2003).
35. Susaki, D., Takeuchi, H., Tsutsui, H., Kurihara, D. & Higashiyama, T. Live imaging and laser disruption reveal the dynamics and cell-cell communication during *Torenia fournieri* female gametophyte development. *Plant Cell Physiol.* **56**, 1031–1041 (2015).
36. Maruyama, D. et al. Independent control by each female gamete prevents the attraction of multiple pollen tubes. *Dev. Cell* **25**, 317–323 (2013).
37. Nagahara, S., Takeuchi, H. & Higashiyama, T. Generation of a homozygous fertilization-defective *gcs1* mutant by heat-inducible removal of a rescue gene. *Plant Reprod.* **28**, 33–46 (2015).
38. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30**, S162–S173 (2009).
39. Miyazaki, S. et al. ANXUR1 and 2, sister genes to FERONIA/SIRENE, are male factors for coordinated fertilization. *Curr. Biol.* **19**, 1327–1331 (2009).
40. Wrzaczek, M. et al. GRIM REAPER peptide binds to receptor kinase PRK5 to trigger cell death in *Arabidopsis*. *EMBO J.* **34**, 55–66 (2015).

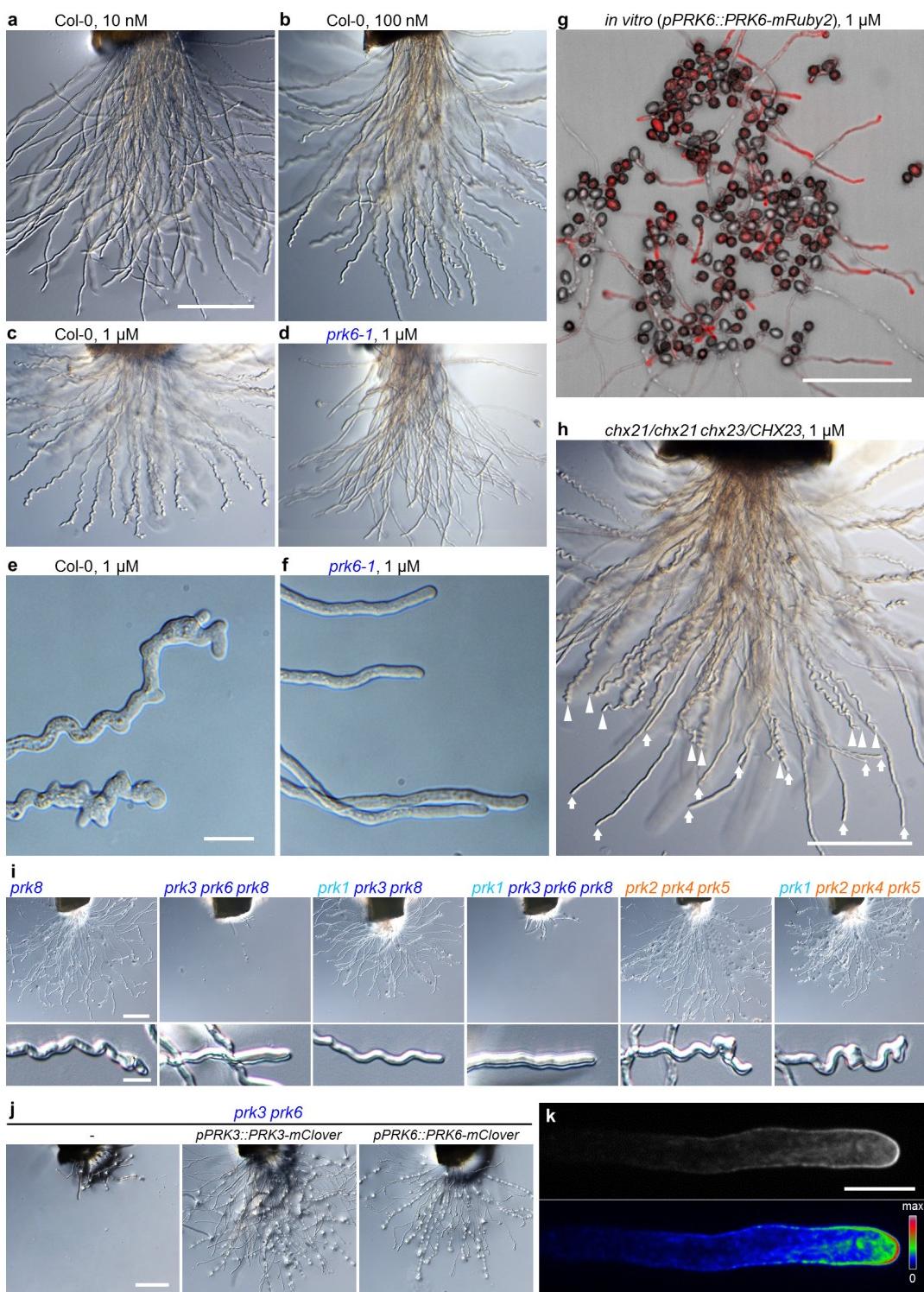


Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Phylogenetic relationship, expression, gene structure and fertility of *A. thaliana* PRK family protein mutants.

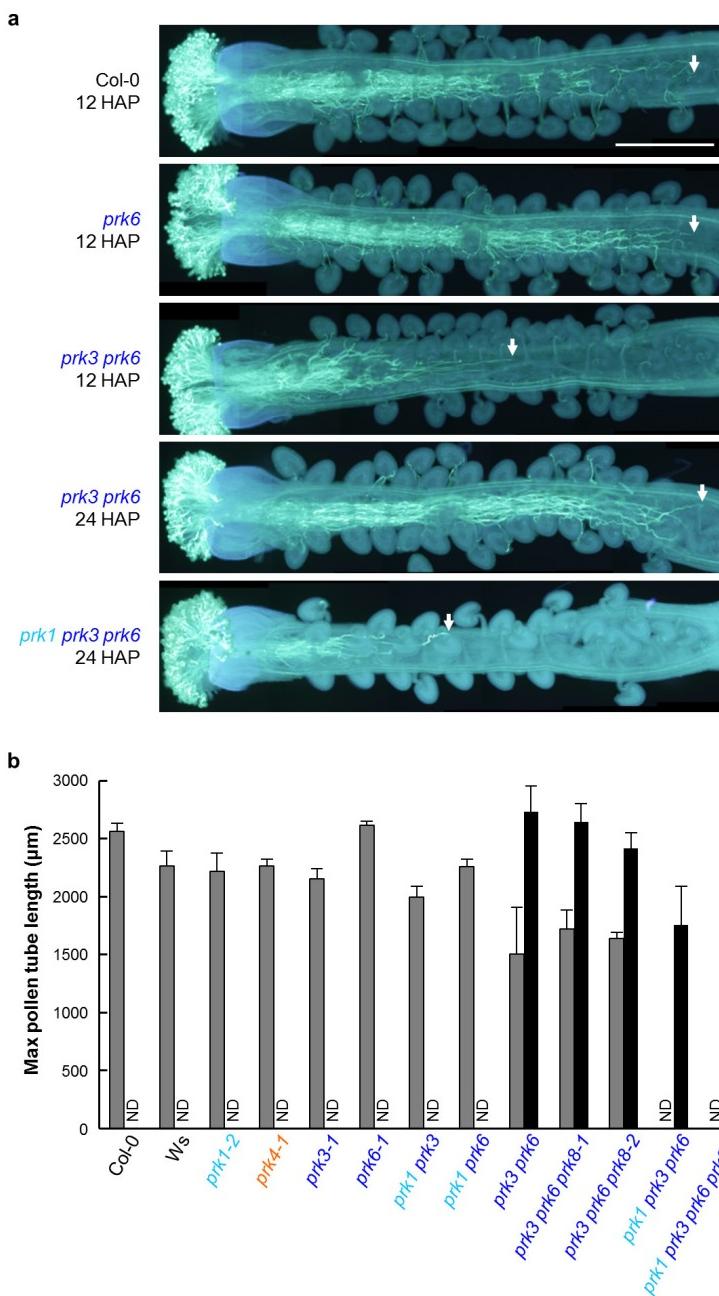
a, A neighbour-joining (NJ) tree constructed using full-length amino-acid sequences of PRK1–PRK6 (ref. 14), PRK7 and PRK8 (assigned in this study). The bootstrap values as percentages and the scale for substitutions per site are shown. **b**, PRK expression during pollen germination and growth. Expression levels are shown using normalized values and standard deviation from microarray data ($n = 4$ for dry pollen, 30 min *in vitro* pollen tube (PT), and 4 h *in vitro* PT; $n = 3$ for semi-*in vivo* PT)¹³. **c**, Structure and T-DNA insertion of PRK genes. Grey boxes show exons, and black boxes show introns or untranslated regions that are registered in The *Arabidopsis* Information Resource (TAIR). The T-DNA insertion sites determined by genomic PCR and sequencing are drawn on the gene structure and indicated in Extended Data Table 1. **d**, Reverse transcription PCR (RT-PCR) analysis of the *prk* single mutants. Anther cDNA was used for the analysis. Positions of the primers are indicated

in the gene structure (c). *ACT2* was used as the loading control. For gel source data, see Supplementary Fig. 1. **e**, **f**, The rate of developing seeds upon self-pollination of *prk* single mutants (e) and upon reciprocal crosses with Col-0 and *prk* multiple mutants (f). Asterisks in e indicate the mutants used for the *prk* multiple mutants in this study. Note that, in addition to multiple mutants of *PRK1* and *PRK3* subclass genes (shown in dark blue), multiple mutants of *PRK1*, *PRK4* and *PRK6*, which are the top three most highly expressed in semi-*in vivo* pollen tubes, and *PRK1*, *PRK2*, *PRK4* and *PRK5*, which form another subclade, were analysed. The *prk1 prk2 prk4 prk5* multiple mutant contains *prk1 prk2 prk5* mutations that cause reduced pollen tube growth *in vitro*¹⁴. Data are mean and s.d. of three (all samples in e; Col-0 pistil \times *prk3-1 prk6-1 pr8-1* and *prk3-1 prk6-1 pr8-2* in f) or four (other samples in f) pistils. **g**, Developing seeds in siliques 8 days after pollination with Col-0, *prk6* and the *prk1 prk3 prk6* triple mutant. The images are representative of four samples. Scale bar, 1 mm.



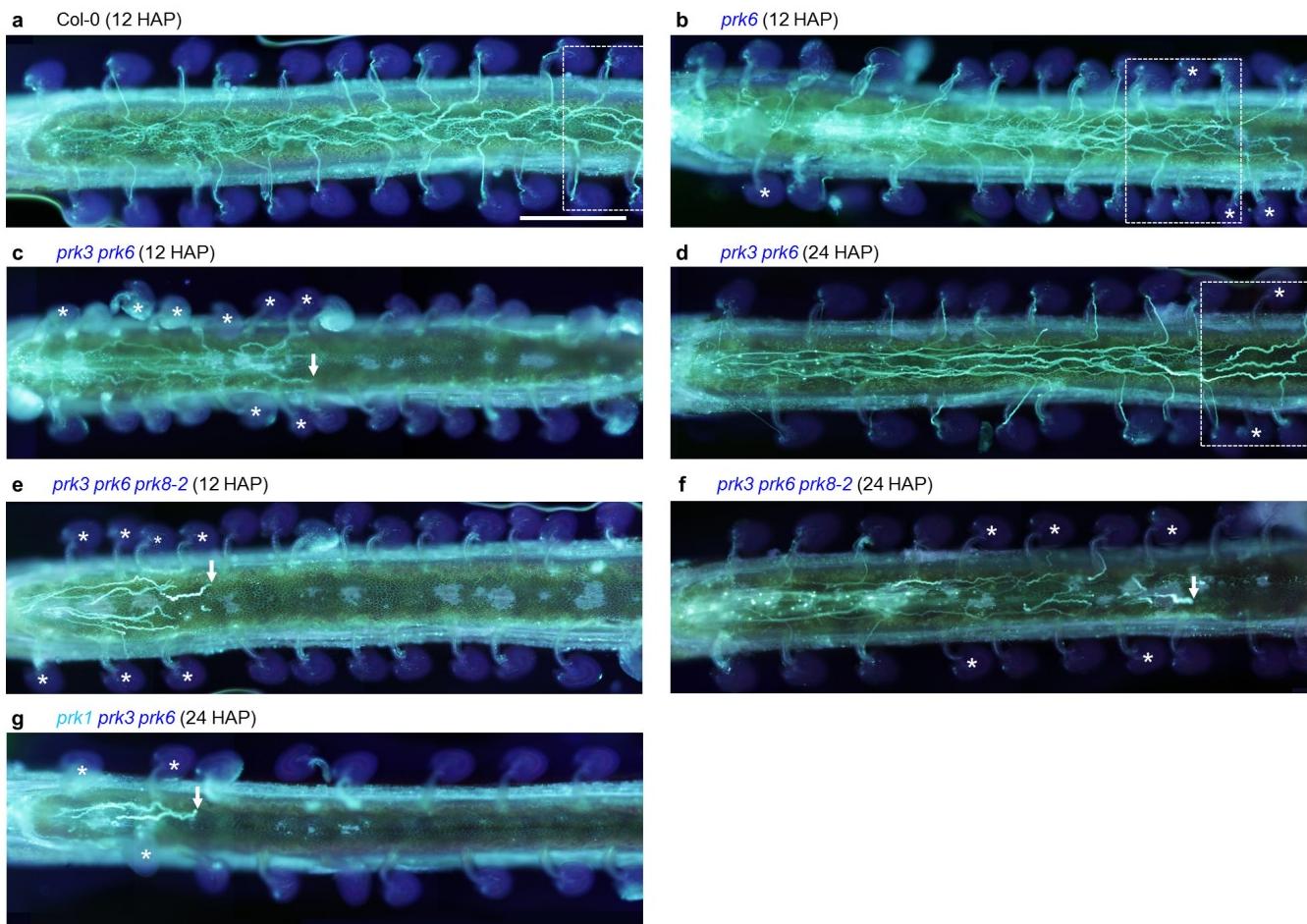
Extended Data Figure 2 | Evaluation of AtLURE1-responsive wavy assay. **a–f**, Semi-*in-vivo* pollen tubes on medium containing the indicated concentrations of AtLURE1.2 peptide. Entire (**a–d**) and magnified (**e, f**) images show wavy and swollen tip growth of wild-type, but not *prk6-1* mutant, pollen tubes in a concentration-dependent manner. Scale bars, 200 μm (**a–d**) and 20 μm (**e, f**). **g**, Growth of PRK6-mRuby2 pollen tubes that directly germinated on medium (that is, *in vitro* pollen tubes) containing 1 μM AtLURE1.2 peptide. **h**, Growth of semi-*in-vivo* pollen tubes from *chx21-1/s1/chx21-1/s1 chx23-4/CHX23* plant¹⁸ on medium containing 1 μM AtLURE1.2 peptide. Roughly half the pollen tubes showed wavy growth as in the wild type (arrowheads), but the rest did not (arrows). These results indicate that *in vitro* pollen tubes and

chx21 chx23 double-mutant pollen tubes have no or less ability to respond to the external AtLURE1 peptide. Scale bars, 200 μm . **i**, Semi-*in-vivo* pollen tube growth and AtLURE1-responsive wavy assay for *prk* mutants additional to those shown in Fig. 1f. Scale bars, 100 μm (top) and 10 μm (bottom). **j**, Complementation of the growth defect in *prk3 prk6* pollen tubes by expression of PRK3-mClover or PRK6-mClover. Note that PRK3-mClover expression restored the growth defect but not the wavy response. The images of **a–j** are representative of at least three assays. **k**, Pollen tube tip localization of PRK3-mClover in a single-plane confocal image (top) and its intensity image by pseudocolour (bottom). The data are representative of three samples. Scale bar, 10 μm .



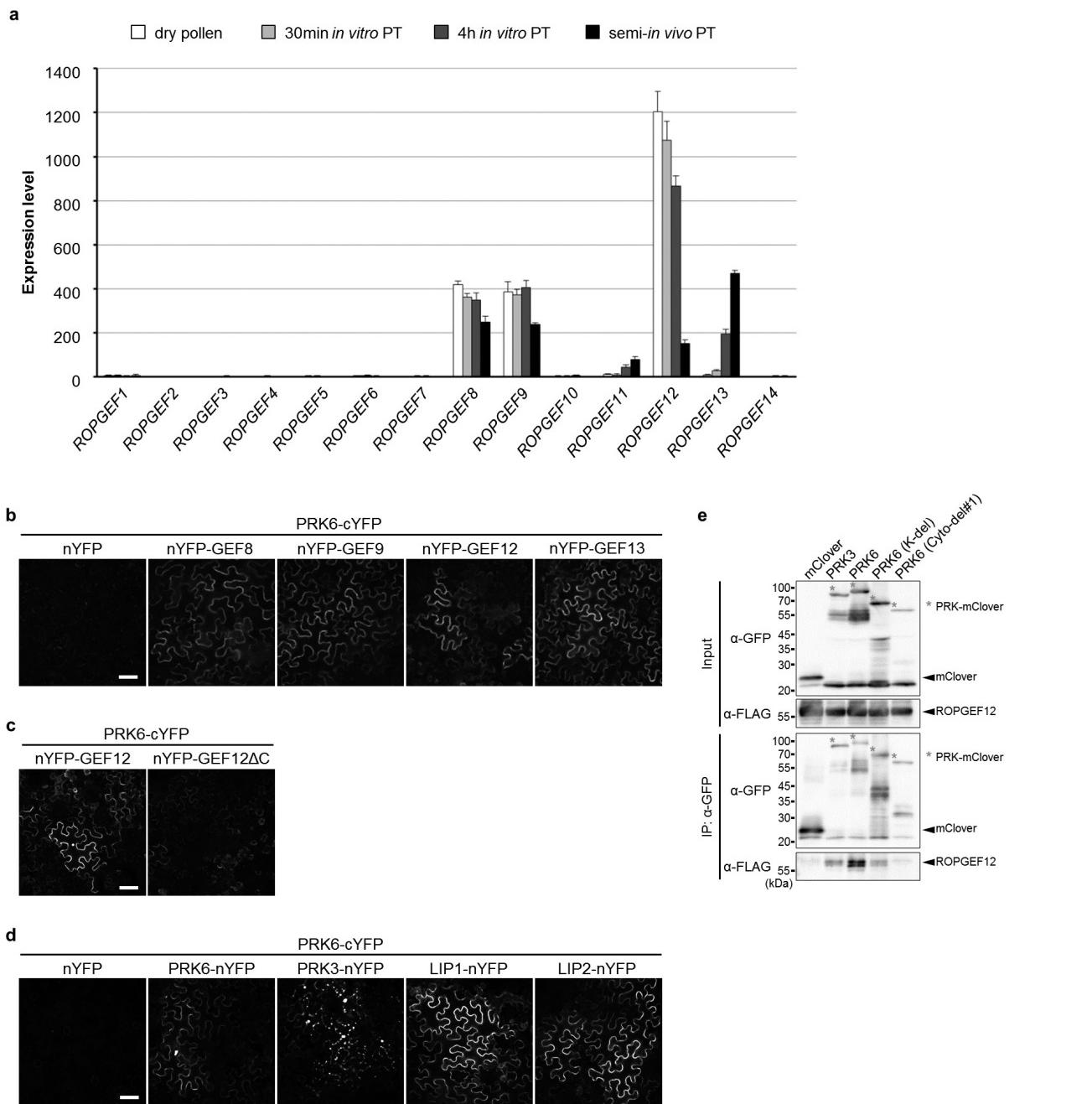
Extended Data Figure 3 | Pollen tube growth of *prk* mutants in the pistil. **a**, Pollen tubes of Col-0, *prk6*, *prk3 prk6* and *prk1 prk3 prk6* growing in the Col-0 pistils. Aniline blue staining was performed 12 or 24 HAP. White arrows indicate the tip of the longest pollen tube in the transmitting tract. Data are representative of three samples for each genotype.

Scale bar, 500 μm. **b**, Length from the top of the stigma to the tip of the longest pollen tube, 12 or 24 HAP with Col-0 and *prk* mutants. About 2,700 μm is the maximum limit for the length in this measurement. The data are the mean and s.d. of three pistils. ND, no data.



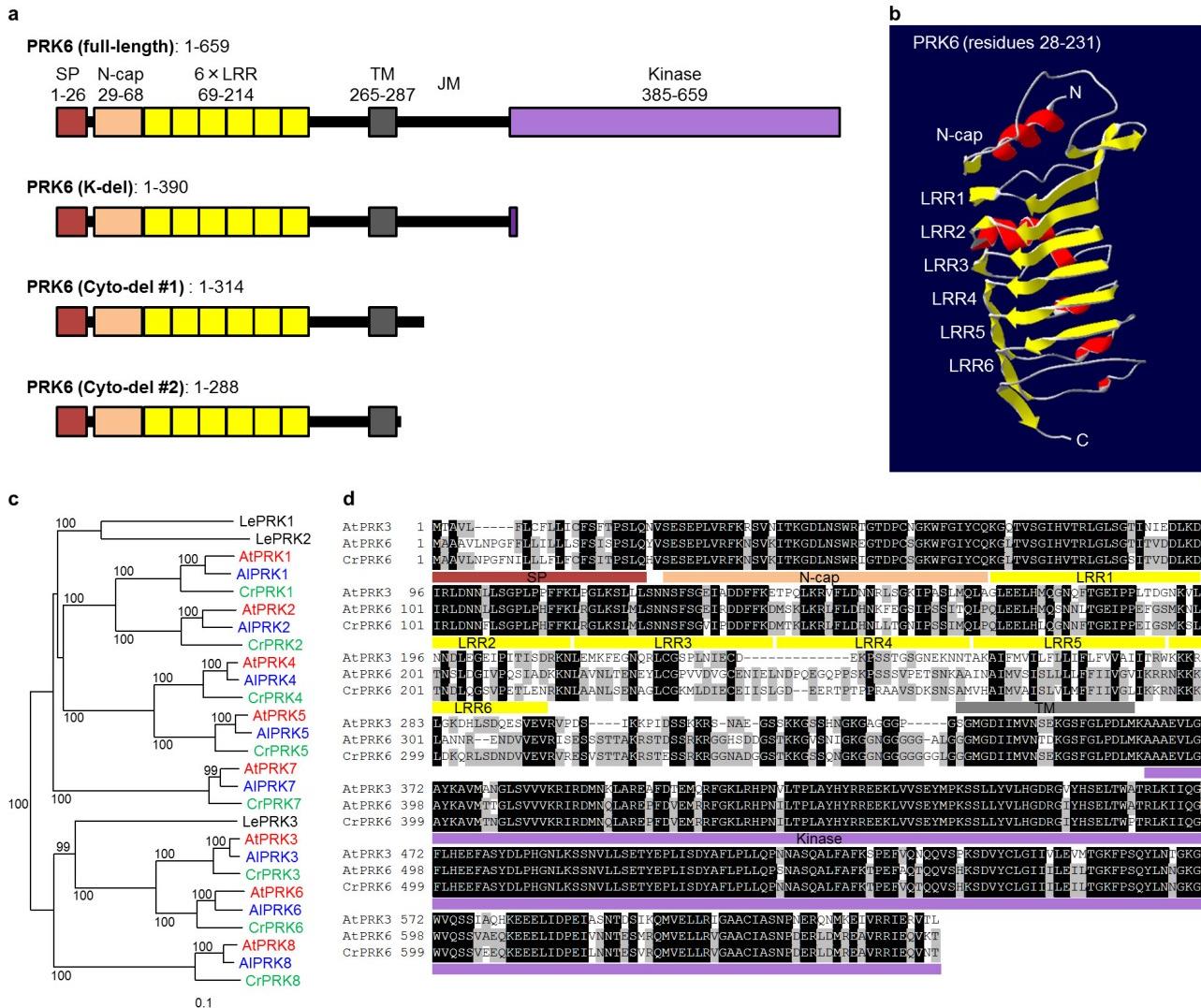
Extended Data Figure 4 | Growth and ovule-targeting of *prk* mutant pollen tubes on the septum surface. a–g, Entire images of growth and ovule-targeting of wild-type (a), *prk6* (b), *prk3 prk6* (c, d), *prk3 prk6 prk8-2* (e, f), and *prk1 prk3 prk6* (g) pollen tubes on the septum surface in wild-type pistils. Arrows indicate the tip of the longest pollen tube on the septum surface. Asterisks mark ovules that did not attract near pollen

tubes. The regions shown in a, b and d are shown in Fig. 2e, f and g, respectively, as higher magnification images. Data are representative of 1–3 images for each genotype. Similar growth properties were observed in a total of 4 samples. Scale bar, 500 µm. Quantitative analysis is shown in Fig. 2d. No analysis was performed for the *prk1 prk3 prk6* mutant because almost no pollen tube reached the ovule.



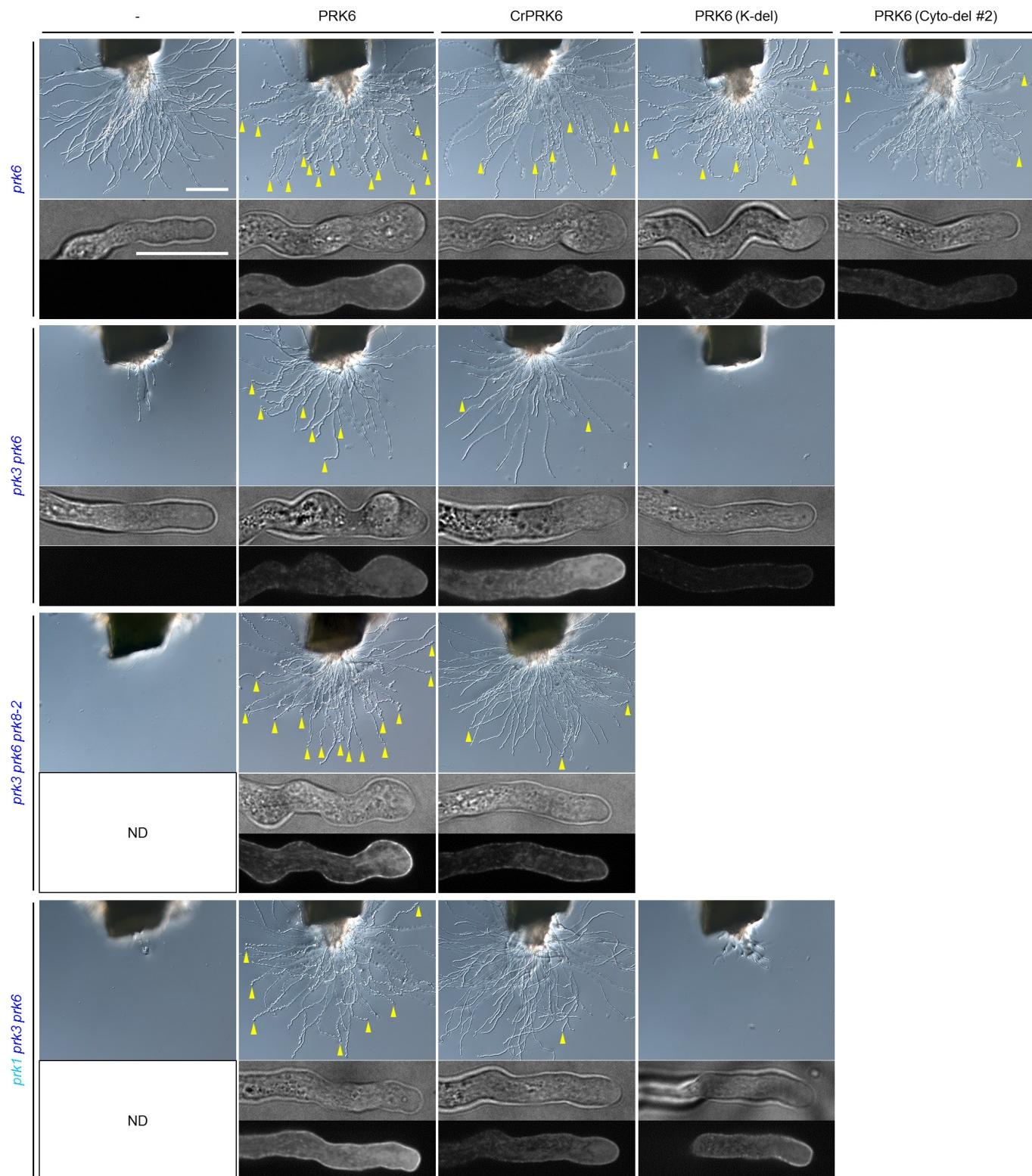
Extended Data Figure 5 | Interaction of PRK6 with pollen-expressed ROPGEFs, PRKs and LIPs. **a**, Gene expression of *ROPGEFs* during pollen germination and growth. The data are normalized expression values and standard deviation from microarray data ($n = 4$ for dry pollen, 30 min *in vitro* PT, and 4 h *in vitro* PT; $n = 3$ for semi-*in vivo* PT)¹³ as noted in Extended Data Fig. 1b. *ROPGEF8*, *ROPGEF9*, *ROPGEF11*, *ROPGEF12* and *ROPGEF13* are expressed specifically in the dry pollen grain and pollen tube. **b**, BiFC assay showing the interaction between PRK6-cYFP and nYFP-GEF8, nYFP-GEF9, nYFP-GEF12, or nYFP-GEF13 (see Methods). **c**, A control experiment using C-terminal-deleted *ROPGEF12*

(*ROPGEF12ΔC*). The C-terminal domain is suggested to mediate the interaction with PRK2 (ref. 8). **d**, BiFC assay showing interaction between PRK6-cYFP and PRK6-nYFP, PRK3-nYFP, LIP1-nYFP or LIP2-nYFP. Scale bars, 50 μ m. Images are representative of more than three experiments. **e**, Co-immunoprecipitation assay of PRK-mClover and ROPGEF12 proteins expressed in *N. benthamiana* leaf cells. ROPGEF12-3 \times Flag protein was precipitated with full-length PRK3, PRK6 and kinase domain-deleted PRK6 (K-del), but not mClover control or cytosolic domain-deleted PRK6 (cyto-del-1). Data are representative of three experiments. For gel source data, see Supplementary Fig. 2.



Extended Data Figure 6 | PRK6 protein structure and PRK proteins of *A. thaliana*, *A. lyrata* and *C. rubella*. **a**, Structures of the PRK6 protein and its deletion version used in this study. The PRK6 extracellular domain contains the N-terminal cap and six LRRs. JM, juxtamembrane domain; N-cap, N-terminal cap; SP, signal peptide; TM, transmembrane domain. The numbers indicate the amino acid ranges of each domain. **b**, A 3D ribbon model of the PRK6 extracellular domain, amino acid residues 28–231, was predicted using the homology modelling platform, SWISS-MODEL (<http://swissmodel.expasy.org/>), and the FLS2 crystal structure (Protein Data Bank (PDB) accession 4MN8) as a template, and was drawn using Swiss-PdbViewer (<http://spdbv.vital-it.ch/>)³⁸. The PRK6 extracellular domain contains the N-terminal cap and six LRRs. **c**, A neighbour-joining tree constructed using PRK protein sequences from tomato (*Lycopersicon esculentum*, LePRK1-3), *A. thaliana*

(AtPRK1–AtPRK8), *A. lyrata* (AlPRK1–AlPRK8), and *C. rubella* (CrPRK1–CrPRK8). The bootstrap values as percentages and the scale for substitutions per site are shown. Accession numbers for *A. lyrata* and *C. rubella* PRKs: AlPRK1 (XP_002868416), AlPRK2 (XP_002883746), AlPRK3 (XP_002877261), AlPRK4 (XP_002883234), AlPRK5 (XP_002891583), AlPRK6 (XP_002871954), AlPRK7 (XP_002867307), CrPRK1 (EOA19015), CrPRK2 (EOA32286, partial sequence), CrPRK3 (EOA25493), CrPRK4 (EOA31871), CrPRK5 (EOA37472), CrPRK6 (EOA23063), CrPRK7 (EOA18255), and CrPRK8 (EOA34527). **d**, A sequence alignment of AtPRK3, AtPRK6 and CrPRK6. Signal peptide, N-terminal cap, LRR1–LRR6, transmembrane domain and kinase domain are indicated beneath the alignment.

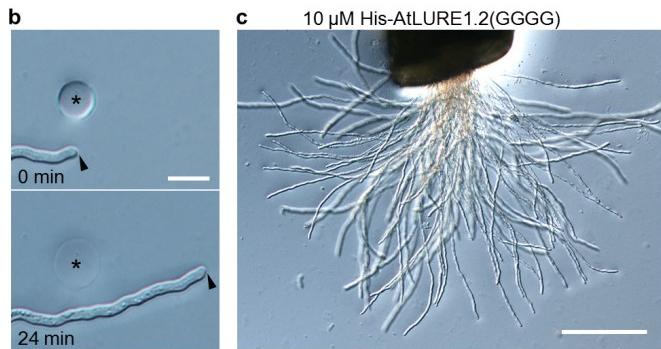


Extended Data Figure 7 | Semi-*in-vivo* pollen tube growth and response to the AtLURE1 peptide of PRK6 variant mutants. Pollen tubes of *prk6*, *prk3 prk6*, *prk3 prk6 prk8-2* and *prk1 prk3 prk6* mutants were assessed in this assay. Full-length PRK6, the PRK6 orthologue of *C. rubella* (CrPRK6), kinase-domain-deleted PRK6 (K-del), and cytosolic-domain-deleted PRK6 (cyto-del-2) were expressed as mRuby2 fusion proteins under the control of their own promoters. Upper differential interference

contrast images show semi-*in-vivo* pollen tube growth in the medium containing the AtLURE1 peptide at 6 HAP. Yellow arrowheads mark some of the pollen tubes showing apparent wavy phenotype. The bottom two images are a bright field image and a confocal image for mRuby2 of a representative pollen tube in the wavy assay. The data are representative images of at least three assays for one or two lines of each genotype. Scale bars, 200 μm (top) and 20 μm (bottom).

a

MKLPIIFITLLIFVSSCTSTLINGSSDEERTYSFSPTTSPFDPRSLNQEL	50
signal peptide	
KIGRIGYC <small>F</small> D <small>C</small> ARACM <small>R</small> RGK <small>Y</small> IRTCSFERKL <small>C</small> RC <small>S</small> ISDIK	90
↓↓ GG G	► AtLURE1.2(GGGG)



Extended Data Figure 8 | A conserved basic amino acid patch of LURE is essential for attraction. **a**, The sequence of full-length AtLURE1.2 accompanied by lysine/arginine residues (yellow highlight) mutated to glycines for AtLURE1.2(GGGG). Cysteine residues in the mature peptide are shown in red. **b, c**, Semi-*in-vivo* attraction assay using gelatine beads containing 5 μ M His-AtLURE1.2(GGGG) (**b**) and wavy assay using 10 μ M His-AtLURE1.2(GGGG) in the medium (**c**). The AtLURE1.2(GGGG) peptide showed no activity in these assays. The data are representative of 14 or 3 samples for **b** or **c**, respectively. Scale bars, 20 μ m (**b**) and 200 μ m (**c**).

Extended Data Table 1 | T-DNA insertion mutants for *PRK* genes

Gene	AGI code	Mutant name*	Mutant ID	Primers (5' to 3')†	Amplified DNA size (bp)‡	Inserted position §	Note	Ref.
<i>PRK1</i>	<i>At5g35390</i>	<i>prk1-1</i>	SALK_112241C	F: aggagtgcacATGCCTCCCATGCAG R: ggatccctcaTTCAAACACCTTGGATC	W: ~680 T: ~690	1st Exon 485(L)/ND	2 primers Hetero↓	14
		<i>prk1-2</i>	SALK_054149	F: aggagtgcacATGCCTCCCATGCAG R: ggatccctcaTTCAAACACCTTGGATC	W: ~680 T: ~410	1st Exon ND /L)451		14, 39
		<i>prk1-3</i>	SALK_074439C	F: ACATCGACGTATCACCAAG R: ggcgcgccATGCAAAGCTGATACTCTC	W: ~410 T: ~360, 460	2nd Exon 1932(L)/(L)1930		
<i>PRK2</i>	<i>At2g07040</i>	<i>prk2-1</i>	SALK_110661	F: aggagtgcacATGGAATCCAAATGTCTC R: agaggcgccCTTGTCCCCCTTCACTTTC	W: ~1000 T: ~430, ~980	1st Exon 225(L)/(L)239	Home or hetero are indistinguishable.	14, 39
<i>PRK3</i>	<i>At3g42880</i>	<i>prk3-1</i>	FLAG_540B08	F: CGAGCAAGAAGCGTAGTAAC R: CCAAGTTAGCTCCGAATG	W: ~540 T: ~450	1st Exon 1088(R)/(L)1117		
<i>PRK4</i>	<i>At3g20190</i>	<i>prk4-1</i>	GK-065H10	F: CCTTAGCAACATGGATCCTG R: CCTCTTCACGACCAGCATC	W: ~610 T: ~450	2nd Exon 880(L)/ND		
<i>PRK5</i>	<i>At1g50610</i>	<i>prk5-1</i>	SALK_016815	F: GGTTTGAGATGTTGCGAAG R: ggcgcgccATCGATTATCGAGAAACCA	W: ~200 T: ~300, ~350	3rd Exon 2291(L)/(L)2286		14, 40
		<i>prk5-2</i>	SALK_101260	F: CAATCCACCAGTTATGCCTG R: CACCTGAGACATTGCCAC	W: ~520 T: ~270, ~620	Promoter -312(L)/(L)-258		39, 40
<i>PRK6</i>	<i>At5g20690</i>	<i>prk6-1</i>	SALK_129244C	F: GGGACCTAAATTATGGAGAG R: TCTGGTACCGAGGAGGATG	W: ~640 T: ~550, ~550	1st Exon 410(L)/(L)419	Two mutant bands are indistinguishable.	
		<i>prk6-2</i>	SALK_076923	F: GGGACCTAAATTATGGAGAG R: TCTGGTACCGAGGAGGATG	W: ~640 T: ~550	1st Exon ND /L)440	2 primers	39
		<i>prk6-3</i>	GK-751H10	F: GGGACCTAAATTATGGAGAG R: TCTGGTACCGAGGAGGATG	W: ~640 T: ~400, ~740	1st Exon 610(L)/(L)657		
<i>PRK7</i>	<i>At4g31250</i>	<i>prk7-1</i>	SALK_105111	F: CCAGTTTACCGGTGAGATAGAC R: CACAGCCCCCTTGTACCTG	W: ~720 T: ~880	1st Exon ND /L)464		
		<i>prk7-2</i>	GK-607A07	F: CCAGTTTACCGGTGAGATAGAC R: CACAGCCCCCTTGTACCTG	W: ~720 T: ~950	1st Intron 1085(L)/ND		
<i>PRK8</i>	<i>At1g72460</i>	<i>prk8-1</i>	SALK_052766C	F: ACGTCGTACCTCTGAACAAG R: GAGAGATTGAGATTGAGCGTC	W: ~790 T: ~290, ~930	1st Intron 923(L)/(L)937	Hetero↓	
		<i>prk8-2</i>	SAIL_1277_H12	F: ACGTCGTACCTCTGAACAAG R: GAGAGATTGAGATTGAGCGTC	W: ~790 T: ~550	2nd Exon 1360(L)/ND		

*Except for *prk1-1*, *prk1-2*, *prk2-1*, *prk5-1* and *prk5-2*, mutant names were assigned in this study.

†Forward (F) and reverse (R) primers were designed in the up- and downstream regions of T-DNA insertion. Lowercase letters are additional nucleotides for the purposes of cloning.

The primers for T-DNA borders are 5'-ATTTGCCGATTTCGGAAC-3' (LBb1.3) for SALK, 5'-AACGCCCCAATGTGTTAAAGTTGTC-3' for SAIL, 5'-CCCATTGGACGTGAATGTAGAC-3' for GABI-Kat (GK), and 5'-CTGATACCGACGTTGCCGCATAA-3' (Tag3) for Flag.

‡Approximate sizes of wild-type (W) and mutant (T) bands, which were amplified by genomic PCR with 3 or 2 primers (for *prk1-1* and *prk6-2*), are shown. Genomic PCR with 3 primers was performed using forward, reverse and T-DNA primers in one reaction. Genomic PCR with 2 primers was performed using forward and reverse primers (for wild-type) and forward or reverse and T-DNA primers (for T-DNA) in two separate reactions.

§The inserted positions were determined by sequencing of genomic PCR products. The numbers indicate genomic nucleotide positions connected to T-DNA or non-genomic sequences. L or R in parentheses shows which border is inserted at the end. ND, not determined (for example, 485(L)/ND means that the four-hundred-and-eight-fifth nucleotide in the genomic sequence is connected to the T-DNA left border, and the junction of another side of the T-DNA was not determined).

||Hetero↓: Heterozygous mutants in these alleles had aborted pollen grains and showed semi-sterility, probably owing to genomic rearrangements, although homozygous mutants in these alleles were normal.

References 14, 39 and 40 are cited in the table.

MIMIVIRE is a defence system in mimivirus that confers resistance to virophage

Anthony Levasseur^{1,2*}, Meriem Bekliz^{1,2*}, Eric Chabrière^{1,2}, Pierre Pontarotti³, Bernard La Scola^{1,2} & Didier Raoult^{1,2}

Since their discovery, giant viruses have revealed several unique features that challenge the conventional definition of a virus, such as their large and complex genomes, their infection by virophages and their presence of transferable short element transpovirons^{1–5}. Here we investigate the sensitivity of mimivirus to virophage infection in a collection of 59 viral strains and demonstrate lineage specificity in the resistance of mimivirus to Zamilon⁶, a unique virophage that can infect lineages B and C of mimivirus but not lineage A. We hypothesized that mimiviruses harbour a defence mechanism resembling the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas system that is widely present in bacteria and archaea^{7–10}. We performed *de novo* sequencing of 45 new mimivirus strains and searched for sequences specific to Zamilon in a total of 60 mimivirus genomes. We found that lineage A strains are resistant to Zamilon and contain the insertion of a repeated Zamilon sequence within an operon, here named the ‘mimivirus virophage resistance element’ (MIMIVIRE). Further analyses of the surrounding sequences showed that this locus is reminiscent of a defence mechanism related to the CRISPR-Cas system. Silencing the repeated sequence and the MIMIVIRE genes restores mimivirus susceptibility to Zamilon. The MIMIVIRE proteins possess the typical functions (nuclease and helicase) involved in the degradation of foreign nucleic acids. The viral defence system, MIMIVIRE, represents a nucleic-acid-based immunity against virophage infection.

Bacteria and archaea acquire immunity to invading genetic elements such as plasmids and phages through the incorporation of short sections of foreign DNA into their genomes⁷. Prokaryotic immunity covers several mechanisms including (1) prevention of viral adsorption and genome injection, (2) cleavage of the invading genome based on the self/non-self-discrimination principle and (3) blockage of phage replication^{8,9}. In terms of prokaryotic immunity, the best characterized models are the restriction-modification (R-M) system and the CRISPR-Cas system^{7,9,11}. The CRISPR system incorporates short fragments of DNA (21–72 nucleotides) and then uses the transcribed RNA as a guide for destroying the invading element⁷. The CRISPR system is therefore able to memorize and discriminately attack the invaders: that is, nucleic acids. The components of the CRISPR-Cas system differ broadly in terms of occurrence, sequence, number of loci and size across bacterial and archaeal genomes. CRISPRs are found in about 48% of bacteria and 80% of archaea, on the basis of the investigation of publicly available genomes¹². The features of the CRISPR-Cas system are determined by Cas proteins, which carry diverse functional domains, such as helicase, nuclease and DNA binding motifs^{8,9}. Thus far, the CRISPR-Cas system has been found in bacteria and archaea^{7,10} and in only one bacteriophage¹³. In this former example, the CRISPR-Cas acquisition is used to counteract a phage inhibitory chromosomal island of the bacterial host, *Vibrio cholerae*¹³. The discovery of giant viruses living together with microbes in an amoeba-filled battlefield

has challenged the traditional definition of a virus^{1–3,14}. mimiviruses are visible with photonic microscopy, have a large and complex genome containing sequences transferred from other organisms¹⁵, can be infected with viral parasites known as virophages and contain transferable short elements that resemble transposons from bacteria^{4,5}. As mimiviruses behave similarly to intra-amoebal microbes^{16,17}, we speculated that they could also harbour several defence mechanisms in the microbial arms race, and specifically searched for a system resembling the CRISPR-Cas system.

Recently, we reported the identification of a novel virophage, Zamilon, which was found to be associated with giant viruses from the *Mimiviridae* family⁶. In the founding members of the family *Mimiviridae*, three lineages, A, B and C, have been identified among the amoebae mimiviruses. Zamilon was able to infect strains of the B (2/2) and C (2/2) lineages of mimivirus but not the two lineage A strains (0/2). Here, we infected with two virophages a collection of 59 *Acanthamoeba polyphaga* mimivirus (APMV) strains, including 28, 8 and 23 strains from the A, B and C lineages, respectively (Extended Data Fig. 1). Two virophages, Sputnik 3 (as positive control) and Zamilon, were selected for analysis and, after 24 h, an increase in Sputnik 3 DNA was observed in all the APMVs (59/59). In contrast, Zamilon was able to replicate in APMV lineages B (8/8) and C (23/23) but not in the strains from lineage A (0/28). These results confirmed and extended our initial observation that all group A strains of mimivirus are resistant to the Zamilon virophage.

As a hallmark of the CRISPR-Cas system, the acquisition of foreign DNA into the CRISPR array is a prerequisite of resistance to foreign genetic elements. Therefore, to identify potential CRISPR-Cas sequences, we performed *de novo* sequencing on 45 mimivirus strains, including lineages A (21 strains), B (5 strains) and C (19 strains). Combining these with 15 APMV genomes that were already available, we then screened all 60 APMV genomes for foreign virophage DNA sequences. A 28-nucleotide-long stretch that was identical to Zamilon DNA was found in all genomes belonging to lineage A (APMV-A) and in one single strain, the *Megavirus chilensis* strain, of the 24 different lineage C genomes (Extended Data Table 1). This sequence is located in open reading frame 4 (ORF4 encoding a protein distantly related to transposase A) of the Zamilon genome (gi|563399744) but absent in Sputnik and is integrated into mimivirus gene R349 and the corresponding orthologous genes in all APMV-A *Mimiviridae*. The RNA predicted from the 28-nucleotide-long stretch of virophage perfectly matched the sequence of the sense strand in all APMV-A excluding the potential formation of RNA duplex. Strikingly, a 15-nucleotide-long sequence derived from this homologous sequence was repeated four times in all APMV-A genomes (28/28) but was not found in group B and C genomes (Extended Data Table 1). There was a significant correlation between Zamilon resistance and presence of the repeated Zamilon sequence in mimiviruses ($P < 0.001$). We therefore suggest that the four 15-nucleotide-long repeated sequences that were

¹Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63, CNRS 7278, IRD 198, INSERM U1095, Marseille, France. ²IHU Méditerranée Infection, Pôle des Maladies Infectieuses, Assistance Publique-Hôpitaux de Marseille, Faculté de Médecine, 27 Boulevard Jean Moulin, 13005 Marseille, France. ³Aix-Marseille Université, CNRS, Centrale Marseille, I2M, UMR7373, FR 4213 - FR Ecocov 3098, équipe EBM, 13331 Marseille, France.

*These authors contributed equally to this work.

exclusively found in all APMV-A genomes are linked to resistance and immunity against Zamilon virophages. We then investigated the chromosomal environment around the repeated insertion, to identify CRISPR-like elements.

We studied the genomic environment for the presence of putative *cas* genes in the vicinity of the four 15-nucleotide repeated sequences found in all the lineage A strains, as identified by bacterial CRISPR. We found a putative phage-type endonuclease (R354) downstream of the four 15-nucleotide repeated sequence locus (Extended Data Table 2). On the basis of structural similarity searches, this protein has been modelled as a lambda exonuclease protein (36% identity), which is a relative of the Cas4 nuclease family¹⁸. Adjacent to the R349 gene containing the inserted Zamilon sequence, we also identified a putative helicase domain associated with a SNF2 domain (ORF R350). This protein contains motifs that are characteristic of the Cas3 protein, which is involved in the type I bacterial CRISPR–Cas system. The R350 SNF2 domain could be involved in a variety of processes including DNA recombination, chromatin unwinding and DNA repair. We also identified a probable RNase III-encoding gene (ORF R343) localized upstream of the repeated sequences (Extended Data Table 2). In bacterial CRISPR, RNase III is responsible for CRISPR-like transcript processing. Additionally, a putative ATP-dependent DNA helicase (L364) was found downstream of the locus (Extended Data Table 2). The putative ATP-dependent DNA helicase has a multi-domain carboxy (C) terminus that includes a conserved domain from superfamily 2 (SF2), a helicase C domain and a DExD domain, as previously described for the Cas3 family.

In summary, the genomic environment in the vicinity of the four 15-nucleotide repeated sequences found in the entire A lineage contains several distant proteins reminiscent of those associated to the CRISPR–Cas system, and these proteins could play a major role in nucleic-acid-based immunity. We propose that this region of the mimivirus genome should be named MIMIVIRE, representing ‘mimivirus virophage resistance element’.

A comparative model between the CRISPR–Cas system and MIMIVIRE is depicted in Extended Data Fig. 2. Important discrepancies exist between the two systems, notably in relation to the sequence-specific recognition of the invading nucleic acids, provided by the derived spacers in prokaryote and by the repeated sequences in MIMIVIRE. Contrary to the prokaryotic system in which the repeats are involved in the structural organization of the CRISPR array, MIMIVIRE is assumed to use the four-time repeated sequence inserted in an open reading frame to provide immunity against Zamilon virophage. These four repeated units appear to be essential for immunity because the presence of only one 15-nucleotide-long unit found in some B and C lineages (inserted in non-orthologous genes) did not confer resistance to Zamilon. In addition, the CRISPR system contains multiple integrated virus-derived spacers and, until now, MIMIVIRE was a priori able to target one virophage from the two known virophage strains. Investigation of forthcoming virophages could help us to unravel the MIMIVIRE system, the generality of the system and, possibly, its adaptive immune mechanism. The occurrence of MIMIVIRE was investigated in each of the APMV strains on the basis of the presence and syntenic organization of potential *cas*-related genes. These genes were conserved in all lineages of APMV-sensitive or -resistant Zamilon virophages, whereas no conservation was found with other *Megavirales* families.

To validate our hypothesis, we systematically investigated the silencing of all potential MIMIVIRE genes in mimivirus by short interfering RNA (siRNA)¹⁹. Consequently, we silenced all genes in the vicinity of the inserted sequence to delimitate and decipher the proteins involved in the MIMIVIRE system. A total of 27 genes were silenced and susceptibility to Zamilon infection was subsequently reported (Fig. 1c). By using quantitative PCR (qPCR), we observed an increased virophage DNA concentration after silencing the gene R354 (encoding the endonuclease), the R350 gene (encoding helicase and

SNF2 domains), and the R349 gene (containing the repeated insert). After 48 h, multiplication of the virophage DNA was 14-fold higher for the R354 gene, 18-fold higher for the R350 gene and 65-fold higher for the R349 gene compared with the control mimivirus (Fig. 1a and Supplementary Table 1). In addition, we also combined silencing of the three MIMIVIRE genes and multiplication of the virophage DNA was 32-fold higher compared with the control. The propagation of the virophages is no higher than the unique silencing of the R349 gene, meaning this gene containing the inserted Zamilon sequences is the central component of the MIMIVIRE system. Additionally, we also demonstrated the propagation of Zamilon virophage particles using transmission electron microscopy (Fig. 1b). No multiplication of the Zamilon virophage was observed following silencing of the other surrounding genes, as confirmed both by qPCR and by transmission electron microscopy. According to these experimental results, we delimitated the MIMIVIRE operon and demonstrated that silencing of three different MIMIVIRE genes could restore mimivirus susceptibility to Zamilon.

Nuclease and helicase activities are known to be central enzymatic functions of the prokaryotic CRISPR–Cas system, in which the Cas3 (type I CRISPR–Cas system) catalyses the unwinding and cleavage of foreign double-stranded DNA (dsDNA) and makes it possible to complete the interference process by destroying the invader nucleic acid. According to our *in silico* inference, the R354 and R350 proteins possess typical nuclease and helicase activities, respectively. To validate the function of the R350 and R354 proteins and to compare the MIMIVIRE system with the CRISPR–Cas model, the two corresponding genes were successfully overexpressed in *Escherichia coli* and the putative nuclease and helicase activities were assayed. Nuclease R354 is assumed to cleave the invading nucleic acid and, as expected, the nuclease activity of the R354 protein was evidenced by unspecific cleavage and partial degradation of dsDNA templates (Extended Data Fig. 3). Moreover, nuclease R354 was more active in the degradation of low GC per cent dsDNA templates (that is, 28–38%) than high GC per cent templates (that is, 50–55%). We found that mimiviruses and virophage (~29%) genes were degraded but not *A. polyphaga* genes (59%) (Extended Data Fig. 3). Consequently, GC per cent cleavage specificity was in total agreement with the MIMIVIRE system immune function against virus propagation, while protecting the host organism. The R350 protein has motifs that are characteristic of helicases (SF2 superfamily) that play a central role in many aspects of the CRISPR-mediated adaptive immune systems. Helicases are known to unwind dsDNA but some helicases can rewind, or anneal, complementary strands of polynucleic acids. The annealing helicases could generate non-specific DNA hybridization and produce chimaeric aggregations of high molecular size. To determine the function of the R354 protein, we used dsDNA templates to study the unwinding/rewinding activities. We systematically observed high molecular aggregates, confirming the biochemical activity of unzipping and zipping the dsDNA, followed by aspecific hybridization of complementary sequences (Extended Data Fig. 3). These high molecular aggregates disappeared after heating and we observed a single band of the expected DNA fragment size that corresponded to the dehydrated molecules.

As demonstrated for the prokaryotic CRISPR system using Cas3 and CASCADE proteins, the helicase–nuclease R350 and nuclease R354 of the MIMIVIRE system confer central enzymatic activities that may be involved in the cleavage of foreign nucleic acid.

Its distant analogy to the bacterial CRISPR–Cas model raises the question of the origin of the MIMIVIRE system. We therefore investigated its evolutionary history by conducting a phylogenetic analysis of the experimentally validated proteins R350 and R354. In APMV, these genes were grouped together and outside their bacterial homologues and other nucleocytoplasmic large DNA viruses (Extended Data Fig. 4). This result suggests that these MIMIVIRE genes were present in the ancestors of these viruses. These two genes could also be found

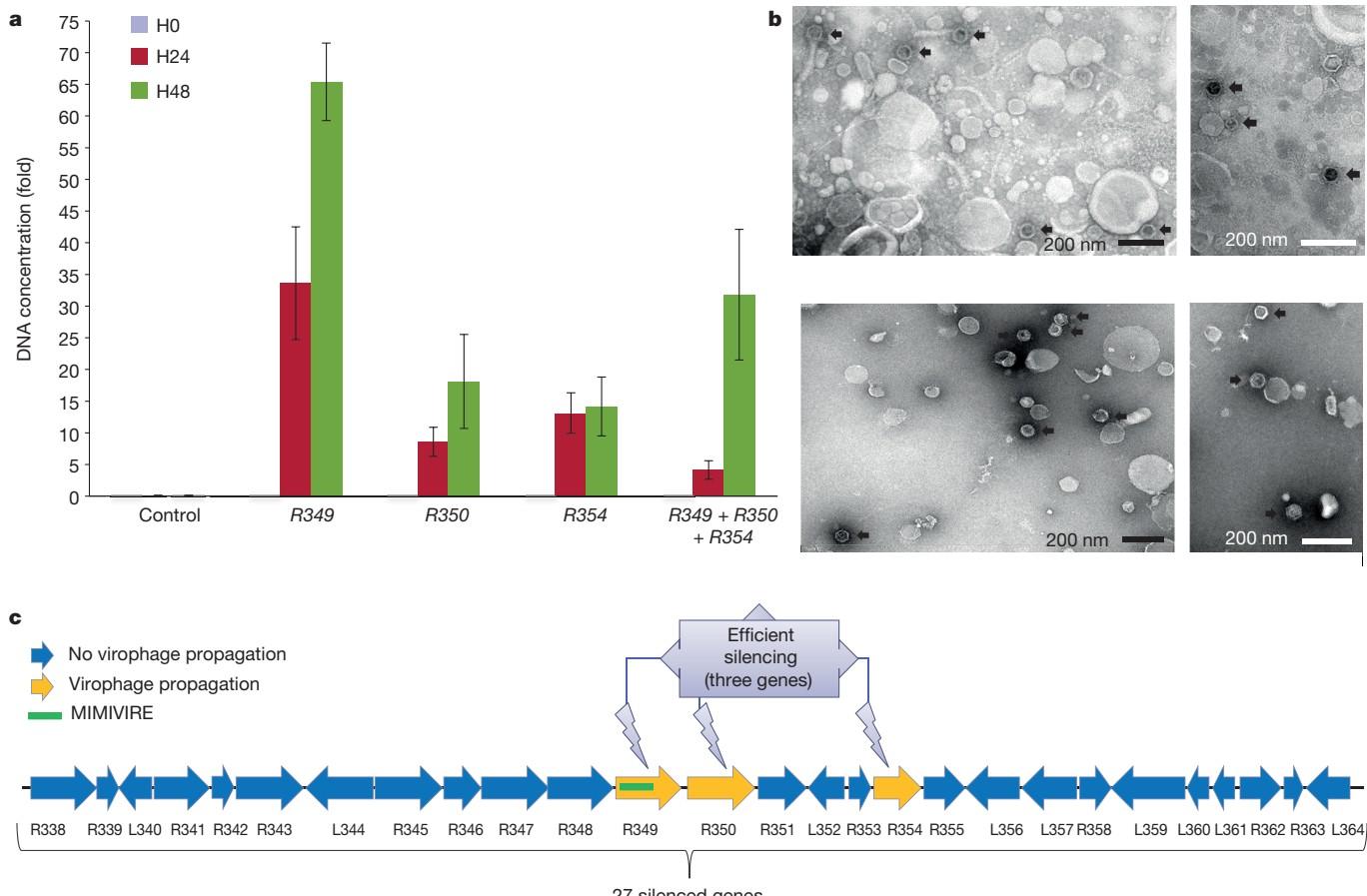


Figure 1 | Chromosomal environment of the MIMIVIRE locus of Mimiviridae and virophage infection. **a**, Quantification of Zamilon propagation after 0 h, 24 h and 48 h (H0, H24 and H48) in the wild-type mimivirus (control) and in the three silenced mimivirus strains (genes R349, R350 and R354). The y axis represents the increase of the DNA

concentration of Zamilon (x -fold) compared with the control. Mean values (\pm s.d.) of three independent experiments. **b**, Negative staining electron microscopy after 48 h of growth; the Zamilon virophage is identified graphically by black arrows. **c**, The 27 silenced genes are indicated with blue (no virophage infection) and yellow (virophage infection) arrows.

in many other viruses, but are scattered along the genome and their role remains to be established. Concerning the R349 gene, no orthologous gene was retrieved in nucleoplasmic large DNA viruses, with the exception of the three APMV lineages.

Giant viruses have extraordinary features that render them unique in the viral world. We therefore tried to identify whether they may also have defence mechanisms similar to those that have been described in bacteria and archaea. We have identified sequences of foreign repeated DNA in these viruses that suggest they have also developed prokaryotic-type defence mechanisms to inhibit the genetic parasitism that they inevitably encounter in their protist hosts²⁰. In this study, we identified a distant CRISPR-Cas-like mechanism called the MIMIVIRE system that explains the resistance of lineage A mimiviruses to the Zamilon virophage. We here unveil this novel immune system in giant viruses, as a result of our computational analysis as previously performed for the initial identification of the CRISPR-Cas system in prokaryotes²¹. We additionally confirmed the biological role of the MIMIVIRE system by silencing and overexpressing two of the genes that are incorporated in it. Both experimental results (silencing of MIMIVIRE genes and functional characterization of MIMIVIRE proteins) confirmed our hypothesis about the fundamental role of MIMIVIRE in the susceptibility of mimivirus to virophage infection and indicated that MIMIVIRE is a defence system against invading elements such as nucleic acids. Besides eliminating competing parasite virophages, MIMIVIRE could also function as a means of maintaining the lytic and infective capacity of the giant virus⁴. In the future, further experimental studies will be required to unravel the molecular bases of the

mechanism that drives the MIMIVIRE system. Our findings illustrate that giant viruses have undergone genetic evolution that is similar to other microbes, via the incorporation of viral parasites (virophages), mobile elements (transpovirons, polintons) and lateral gene transfer²², and that MIMIVIRE confers a nucleic-acid-based immunity in giant viruses.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 April 2015; accepted 18 January 2016.

Published online 29 February 2016.

- La Scola, B. et al. A giant virus in amoebae. *Science* **299**, 2033 (2003).
- Koonin, E. V. & Dolja, V. V. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Mol. Biol. Rep.* **78**, 278–303 (2014).
- Forterre, P. Giant viruses: conflicts in revisiting the virus concept. *Intervirology* **53**, 362–378 (2010).
- La Scola, B. et al. The virophage as a unique parasite of the giant mimivirus. *Nature* **455**, 100–104 (2008).
- Raoult, D. How the virophage compels the need to readdress the classification of microbes. *Virology* **477**, 119–124 (2015).
- Gaia, M. et al. Zamilon, a novel virophage with Mimiviridae host specificity. *PLoS ONE* **9**, e94923 (2014).
- Makarova, K. S. et al. An updated evolutionary classification of CRISPR-Cas systems. *Nature Rev. Microbiol.* **13**, 722–736 (2015).
- Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* **41**, 4360–4377 (2013).
- Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).

10. Barrangou, R. & Marraffini, L. A. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54**, 234–244 (2014).
11. Bickle, T. A. & Kruger, D. H. Biology of DNA restriction. *Microbiol. Rev.* **57**, 434–450 (1993).
12. Pourcel, C. & Drevet, C. in *CRISPR-Cas Systems* 33–59 (Springer, 2012).
13. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–491 (2013).
14. Raoult, D. TRUC or the need for a new microbial classification. *Intervirology* **56**, 349–353 (2013).
15. Filée, J. & Chandler, M. Gene exchange and the origin of giant viruses. *Intervirology* **53**, 354–361 (2010).
16. Boyer, M. et al. Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc. Natl Acad. Sci. USA* **108**, 10296–10301 (2011).
17. Raoult, D. & Boyer, M. Amoebae as genitors and reservoirs of giant viruses. *Intervirology* **53**, 321–329 (2010).
18. Zhang, J., Kasciukovic, T. & White, M. F. The CRISPR associated protein Cas4 is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PLoS ONE* **7**, e47232 (2012).
19. Sobhy, H., La Scola, B., Pagnier, I., Raoult, D. & Colson, P. Identification of giant Mimivirus protein functions using RNA interference. *Front. Microbiol.* **6**, 345 (2015).
20. Sliamni, M., Pagnier, I., Raoult, D. & La Scola, B. Amoebae as battlefields for bacteria, giant viruses, and virophages. *J. Virol.* **87**, 4783–4785 (2013).
21. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAs, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7 (2006).
22. Yutin, N., Raoult, D. & Koonin, E. V. Virophages, polintons, and transpovirons: a complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. *Virol. J.* **10**, 158 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Benamar, A. Caputo and C. Robert for technical support, and L. Pinault for her assistance on protein production.

Author Contributions A.L., B.L. and D.R. conceived the project and designed the study and experiments; A.L., M.B., P.P., B.L. and D.R. analysed the results; A.L. and D.R. wrote the manuscript. All authors read and approved the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R. (didier.raoult@gmail.com) or B.L. (bernard.la-scola@univ-amu.fr).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Genome sequencing and bioinformatic analyses. Forty-five new *Mimiviridae* strains were isolated and subsequently sequenced using either a 454-Roche GS FLX Titanium system (Roche Diagnostics), AB SOLiD instrument (Life Technologies) or MiSeq sequencer (Illumina) (Supplementary Table 2). In detail, 24 *Mimiviridae* strains were sequenced on the MiSeq Technology (Illumina) with a paired-end and barcode strategy on different flowcells using a Nextera XT library kit (Illumina). The DNA was quantified by a Qubit assay with a high sensitivity kit (Life Technologies) and dilution was performed to require 1 ng of each sample as input. The construction of the library was performed by a 'fragmentation' step to fragment the genomic DNA, followed by limited cycle PCR amplification to complete the tag adapters and introduce dual-index barcodes. Automated cluster generation and paired-end sequencing was performed on a MiSeq instrument in a single 39-h run to 2 × 250 bp. The sequencing strategies of 23 *Mimiviridae* strains were performed through the SOLiD 4_Life technologies in NGS technologies. The paired-end library was constructed from 1 µg of purified genomic DNA of each strain. The sequencing was performed to 50 × 35 bp using SOLiD V4 chemistry on one full slide on an Applied Biosystems SOLiD 4 machine. All of these 96 genomic DNAs were barcoded with the module 1-96 barcodes provided by Life Technologies. Thirteen strains of the *Mimiviridae* paired-end library were pyrosequenced on the 454 Roche Titanium. Each project was loaded on a 1/4 region on PTP Picotiterplate. The library was constructed with 5 µg of DNA according to the 454 Titanium paired-end protocol and the manufacturer's instructions. It was mechanically fragmented on a Covaris device (KBioScience-LGC Genomics) through a miniTUBE-Red 5 kb. The library was clonally amplified in emPCR reactions with a GS Titanium SV emPCR Kit (Lib-L) version 2, then loaded on a GS Titanium PicoTiterPlates PTP Kit 70×75 sequenced with a GS Titanium Sequencing Kit XLR70 and reads generated with an average of 280 bp.

Genome assembly and structural annotation. The Newbler assembler version 2.7 and Abyss genomics version 2.3 assembler were used to assemble *Mimiviridae* genomes (Supplementary Table 2). SOLiD reads were mapped on assembled *Mimiviridae* genome using the CLC Genomics Workbench version 7.5. Gene predictions were performed using GeneMarkS software with default parameters²³.

Virophage DNA screening in APMV. The genomes of Zamilon (NC_022990), Sputnik 1 (EU606015), Sputnik 2 (NC_023846) and Sputnik 3 (NC_023847) were downloaded from the National Center for Biotechnology Information (NCBI). The genomes were fragmented into short fragments of 40 nucleotides using a sliding window of size 10 nucleotides. All fragments were blasted against the respective APMV genomes using an *e*-value threshold = e^{-3} . We then looked for all fragments (with 100% identity) present in the entire lineage A of APMV and mostly absent in lineage B and C. One hit, 28 nucleotides in length, fulfilled these criteria and was selected.

Phylogenetic tree construction. From each query sequence, a data set of putative homologous sequences was built by a BLAST²⁴ run on the NCBI non-redundant (NR) database. The raw data set was manually filtered to eliminate potentially non-homologous sequences, disturbing alignments and duplicates. Alignments were conducted using MUSCLE²⁵. For phylogenetic reconstruction, we used the maximum likelihood method.

APMV and virophage production. The *A. polyphaga* Link-AP1 trophozoite strain²⁶ was cultured in peptone-yeast extract glucose (PYG) medium at 32 °C for 3 days, as described previously²⁷. The giant viruses in our collection were co-cultured with fresh *A. polyphaga* in PYG medium. To purify the giant viruses, the co-culture was centrifuged at low speed (1,700 g per 10 min), and the supernatant was filtered across a 0.8 µm membrane to remove residual amoebas and cysts. Each supernatant was then washed three times with Page's modified Neff's amoeba saline (PAS) by centrifugation at high speed (10,300 g per 10 min) to pellet the virus. Sputnik 3 and Zamilon virophages were produced in co-culture with Mamavirus and Mont1, respectively, in PYG medium containing the amoeba *A. polyphaga*. After complete lysis, the supernatant that was obtained following centrifugation at high speed (10,300 g per 10 min) was successively filtered with 0.8 µm, 0.45 µm and 0.22 µm membranes to obtain a pure virophage suspension. A final ultracentrifugation was performed at 13,900 g for 1.5 h to concentrate each virophage filtrate.

Virophage co-culture with Mimiviridae. *A. polyphaga* were suspended three times in PAS. One million APMV virions were inoculated individually into 10 ml of 5 × 10⁵ cells per millilitre of rinsed *A. polyphaga* that contained 100 µl of either Sputnik 3 or Zamilon suspension. The co-culture was incubated for 1 h at 32 °C, and the supernatant was delicately removed to purge the

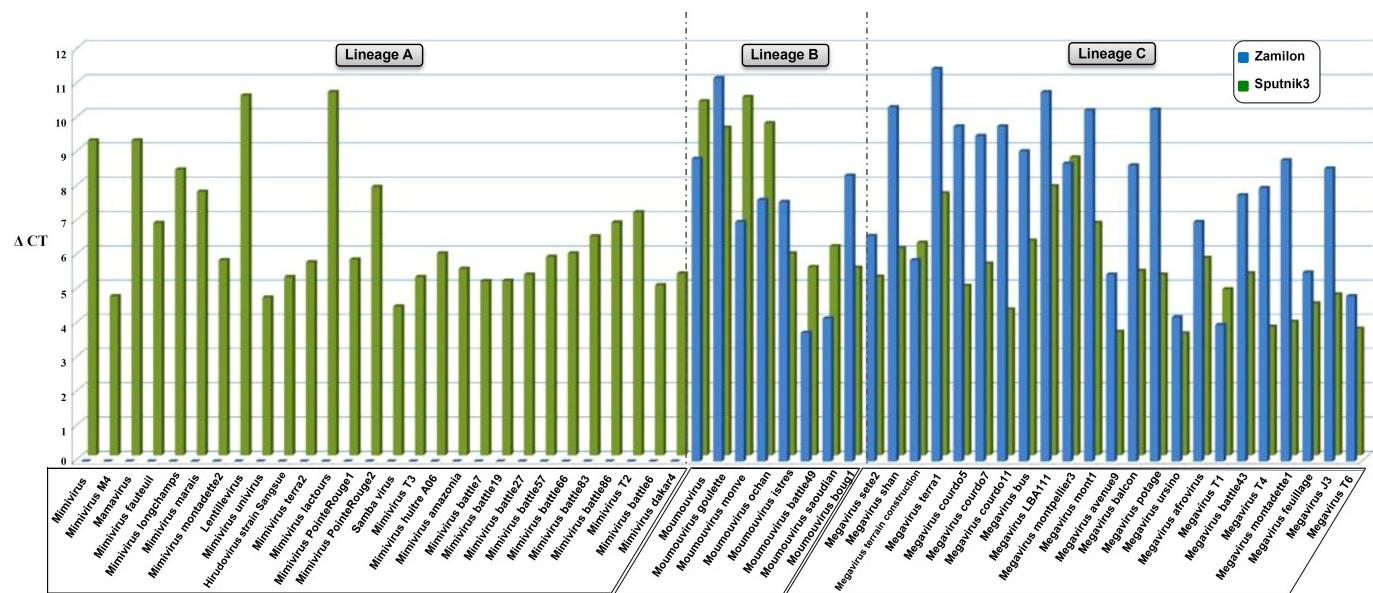
virophage and APMV particles that did not enter the amoebas. Following this procedure, 10 ml of fresh PAS was added. This time point was defined as H0. Each virus was separately incubated without virophage to serve as a negative control. Mamavirus and Mont1 virus that were naturally infected with Sputnik 3 and Zamilon, respectively, were used as positive controls. At 0 and 24 h after infection, a 200 µl aliquot of co-culture was removed for DNA extraction and qPCR to enable the evaluation of virophage multiplication. The DNA extraction was performed using an EZ1 DNA Tissue Kit (Qiagen) according to the manufacturer's instructions. The qPCRs were performed in a CFX96 thermal cycler (Bio-Rad Laboratories) using a SYBR Green PCR Master Mix (Qiagen). Virophages were detected and quantified using primers targeting ORF20 for Sputnik 3 (forward primer 5'-GAGATGCTGATGGAGCCAAT-3', reverse primer 5'-CATCCCACAAGAAAGGAGGA-3') and ORF06 for Zamilon (forward primer 5'-GGGATGAACATCAAGCTGGT-3', reverse primer 5'-GGGTGTTGGAAGCTGACAT-3').

Co-culture and mimivirus silencing. We targeted the mimivirus operon genes using siRNA, an oligonucleotide primer system, which was purchased from Invitrogen (<http://rnadesigner.invitrogen.com/rnaiexpress/>) (Supplementary Table 3). We diluted each 20 µM solution of duplex siRNA and 50 µl of Lipofectamine RNAiMAX (Invitrogen) in 200 µl PAS according to the manufacturer's instructions and recommendations. To improve siRNA specificity, we used duplex siRNA and checked for specific and non-specific pairing. One hour before transfection, 1.5 × 10⁶ *A. polyphaga* were put onto a plate with 5 ml of PAS to allow them time to adhere. After this, the siRNA-Lipofectamine suspension, 10⁶ mimivirus particles and 10¹⁰ Zamilon virophage particles were all added to the plate containing the amoeba. The co-culture was incubated for 1 h at 32 °C, then the supernatant was delicately removed after centrifugation (1,700 g per minute), to eliminate the mimivirus and Zamilon particles that did not enter into the amoebas. The supernatant was replaced by 5 ml of fresh PAS containing the original concentration of siRNA-Lipofectamine, and the culture was submitted for a second incubation for 24–48 h at 32 °C. This time point was defined as H0. The same procedure was used with the omission of Zamilon and/or of mimivirus to serve as negative controls. To control siRNA transfection inside amoeba, a DMI6000 (Leica DMI 6000B) fluorescence microscope was used to visualize the green fluorescence of the oligonucleotides that were transfected into the amoeba. At 0, 24 and 48 h after infection, a 200 µl aliquot of co-culture was removed for DNA extraction and for real-time qPCR to evaluate Zamilon virophage multiplication. Twofold serial dilutions of Zamilon DNA from virophages that were cultivated either with mimivirus (wild type or silenced strains) or with *Mimiviridae* lineage B (Moumouvirus) and lineage C (Courdo7) strains were subjected to qPCR. The Zamilon DNA concentration was subsequently estimated at 0, 24 and 48 h for each condition. For the co-silencing of several genes of mimivirus, we used the same procedure previously mentioned according to the manufacturer's instructions and recommendations.

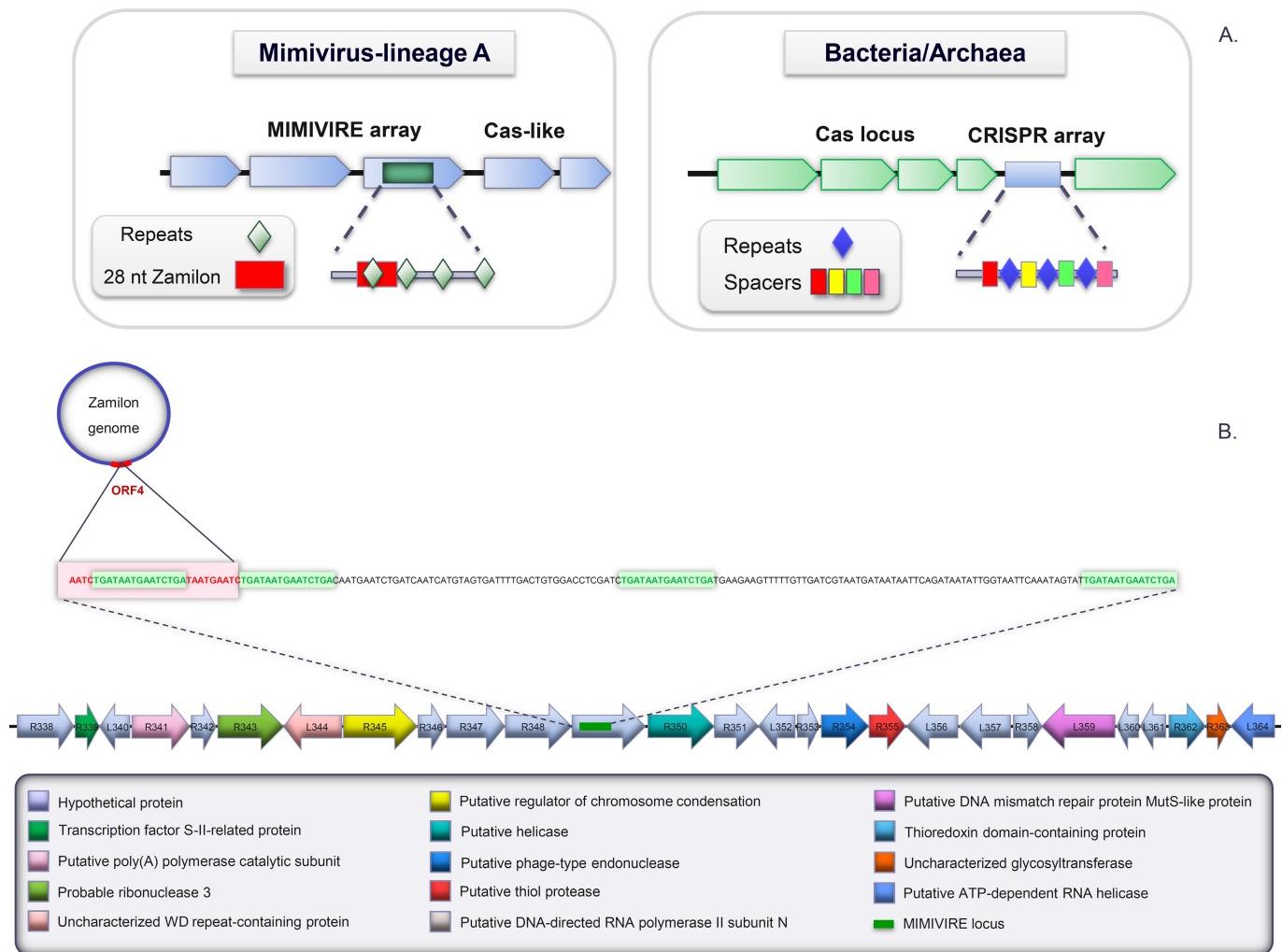
Cloning, expression and purification. Genes encoding the proteins R350 and R354 from APMV were codon-optimized for *E. coli* expression and synthesized by GenScript. Those optimized genes were designed to include a polyhistidine tag at the amino (N) terminus of each protein. Each gene was inserted between the NdeI and NotI cutting sites of a pET22b(+) plasmid. Recombinant proteins were expressed in *E. coli* BL21(DE3)-pGro7/GroEL (TaKaRa) using ZYP-5052 media. Each culture was grown at 37 °C until reaching an absorbance at 600 nm of 0.8 followed by addition of L-arabinose (0.2% m/v) and induction with a temperature transition to 18 °C over 20 h. Cells were harvested by centrifugation (4,250 g, 30 min, 4 °C) and the resulting pellets were resuspended in wash buffer (50 mM Tris pH 8, 300 mM NaCl, 10 mM imidazole) and stored at -80 °C overnight. Frozen cells were thawed and incubated on ice for 1 h after adding lysozyme, DNase I and phenylmethylsulfonyl fluoride (PMSF) to final concentrations of, respectively, 0.25 mg ml⁻¹, 10 µg ml⁻¹ and 0.1 mM. Partly lysed cells were then disrupted by three consecutive cycles of sonication (30 s, amplitude 45) performed on a Q700 sonicator system (QSonica). Cellular debris was discarded after a centrifugation step (21,640 g, 20 min, 4 °C). The recombinant proteins were purified using immobilized metal affinity chromatography (wash buffer: 50 mM Tris pH 8, 300 mM NaCl, 10 mM imidazole; elution buffer: 50 mM Tris pH 8, 300 mM NaCl, 500 mM imidazole) on a 5 ml HisTrap FF crude column (GE Healthcare). Fractions containing each protein of interest were pooled and further purified using size-exclusion chromatography (buffer: 50 mM Tris pH 8, 300 mM NaCl) on a Superdex 75 16/60 column (GE Healthcare). Protein purity was assessed using 10% SDS-PAGE analysis (Coomassie stain). Bands matching the masses of the two proteins of interest were submitted to mass spectrometry analysis, which confirmed the expression of both desired proteins. Protein concentrations were measured using a Nanodrop 2000c spectrophotometer (Thermo Scientific).

Enzymatic treatments. Enzymatic reactions were performed by incubating each PCR product in combination with one of the enzymes (nuclease R354 or helicase–nuclease R350) or both enzymes together. The enzymatic reactions were conducted in PAS buffer solution at 32 °C for 2 h, using a protein concentration of 0.5 mg ml⁻¹ for each enzyme. After incubation, the material was loaded onto agarose gel electrophoresis (1.5%). The DNA products used are listed in Extended Data Fig. 3. Controls were performed with different treatment parameters, such as the denaturation of the enzyme by heating at 94 °C for 10 min, denaturation of the enzyme coupling with DNA by heating at 94 °C for 10 min, 2 h after incubation, denaturation of DNA product by heating at 94 °C for 10 min before coupling with an enzyme and treatment time.

23. Besemer, J., Lomsadz, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618 (2001).
24. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
25. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
26. La Scola, B., Mezi, L., Weiller, P. J. & Raoult, D. Isolation of *Legionella anisa* using an amoebic coculture procedure. *J. Clin. Microbiol.* **39**, 365–366 (2001).
27. Mba Medie, F., Ben Salah, I., Henrissat, B., Raoult, R. & Drancourt, M. *Mycobacterium tuberculosis* complex mycobacteria as amoeba-resistant organisms. *PLoS ONE* **6**, e20499 (2011).

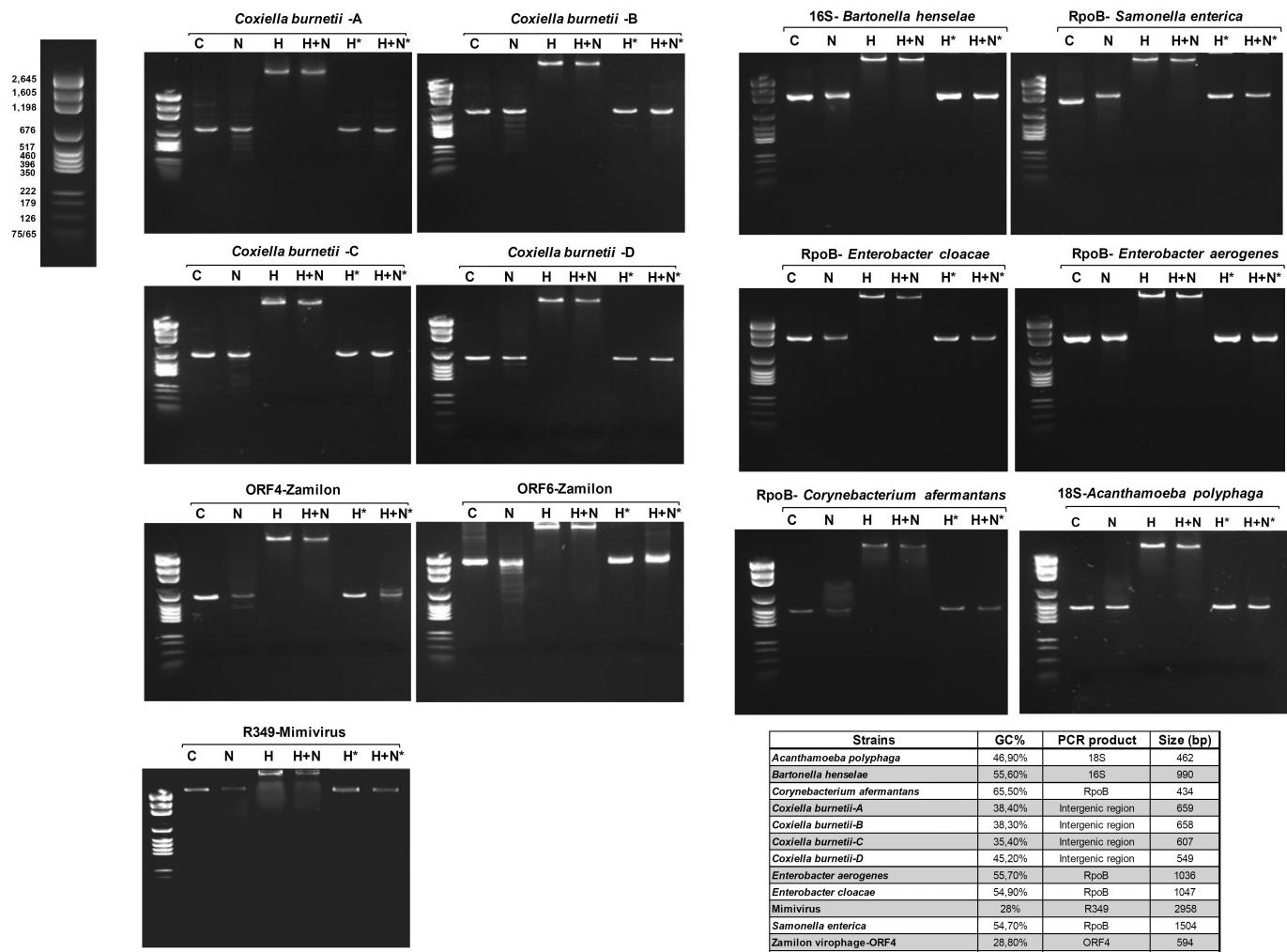


Extended Data Figure 1 | Histogram depicting the replication of Zamilon and Sputnik 3 DNA in Mimiviridae after its phylogenetic classification into lineages A, B and C. The replication of each virophage was measured after 24 h using qPCR. The term Δ CT corresponds to the difference between the CT value specific to virophage at H0 and H24.


Extended Data Figure 2 | The MIMIVIRE defence system.

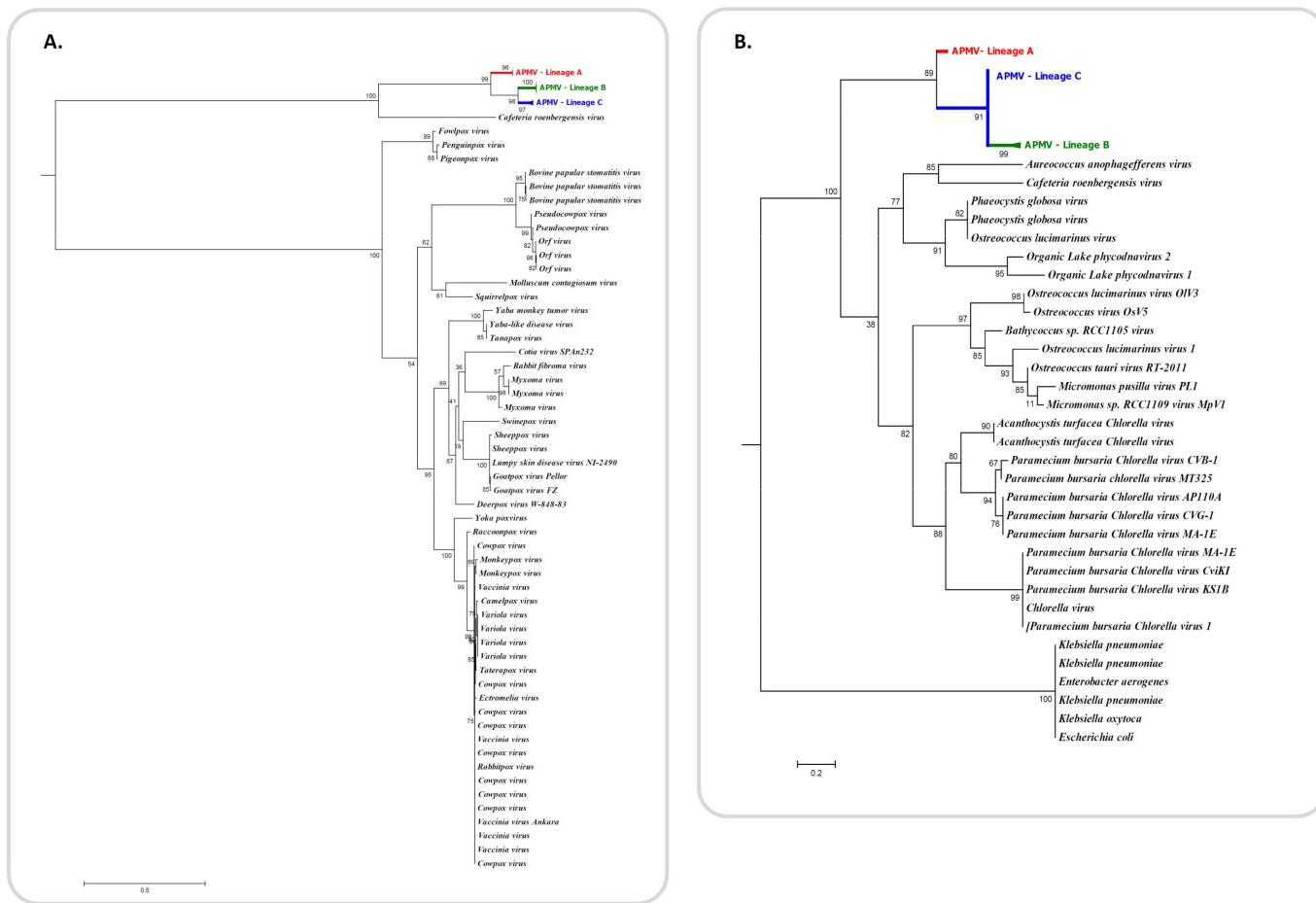
a, A comparative model between prokaryotic CRISPR–Cas system and the viral MIMIVIRE system in APMV-A. **b**, The chromosomal environment of *Mimiviridae* lineage A is illustrated using mimivirus as an example. This organization is conserved across all APMV-A

genomes. The 28-nucleotide-long Zamilon insert sequence is AATCTGATAATGAATCTGATAATGAATC, and the derived 15-nucleotide repeated unit is TGATAATGAATCTGA. The four repeats units are separated by 9, 48 and 63 nucleotides, respectively.



Strains	GC%	PCR product	Size (bp)
<i>Acanthamoeba polyphaga</i>	46.90%	18S	462
<i>Bartonella henselae</i>	55.60%	16S	990
<i>Corynebacterium afermantans</i>	65.50%	RpoB	434
<i>Coxiella burnetii</i> -A	38.40%	Intergenic region	659
<i>Coxiella burnetii</i> -B	38.30%	Intergenic region	658
<i>Coxiella burnetii</i> -C	35.40%	Intergenic region	607
<i>Coxiella burnetii</i> -D	45.20%	Intergenic region	549
<i>Enterobacter aerogenes</i>	55.70%	RpoB	1036
<i>Enterobacter cloacae</i>	54.90%	RpoB	1047
Mimivirus	28%	R349	2958
<i>Samonella enterica</i>	54.70%	RpoB	1504
Zamilion virophage-ORF4	28.80%	ORF4	594
Zamilion virophage-ORF6	35.30%	ORF6	1830

Extended Data Figure 3 | Agarose gel electrophoresis of different DNA products treated with and without nuclease and/or helicase enzymes. C, control; N, nuclease treatment for 2 h; H, helicase treatment for 2 h; H+N, helicase and nuclease treatment for 2 h; H*, helicase treatment for 2 h followed by heating at 94 °C for 10 min; H+N*, helicase and nuclease treatment for 2 h followed by heating at 94 °C for 10 min.



Extended Data Figure 4 | Phylogenetic trees based on the sequences of the two Cas proteins. **a**, R350. **b**, R354.

Extended Data Table 1 | Identification of the Zamilon sequences that were found inserted into the 60 genomes of Mimiviridae

Mimiviridae	Lineage	ZAMILON inserted sequence (28 bp length)	ZAMILON repeat CRISPR-like hits (15 bp length)
Mimivirus	A	1	4
Mimivirus M4	A	1	4
Mamavirus	A	1	4
Mimivirus fauteuil	A	1	4
Mimivirus longchamps	A	1	4
Mimivirus marais	A	1	4
Mimivirus montadette2	A	1	4
Lentillevirus	A	1	4
Mimivirus univirus	A	1	4
Hirudovirus strain Sangsue	A	1	4
Mimivirus terra2	A	1	4
Mimivirus lactours	A	1	4
Mimivirus PointeRouge1	A	1	4
Mimivirus PointeRouge2	A	1	4
Samba virus	A	1	4
Mimivirus T3	A	1	4
Mimivirus huitre A06	A	1	4
Mimivirus amazonia	A	1	4
Mimivirus battle7	A	1	4
Mimivirus battle19	A	1	4
Mimivirus battle27	A	1	4
Mimivirus battle57	A	1	4
Mimivirus battle66	A	1	4
Mimivirus battle83	A	1	4
Mimivirus battle86	A	1	4
Mimivirus T2	A	1	4
Mimivirus battle6	A	1	4
Mimivirus dakar4	A	1	4
Moumouvirus	B	0	1
Moumouvirus goulette	B	0	0
Moumouvirus monve	B	0	1
Moumouvirus ochan	B	0	0
Moumouvirus istres	B	0	0
Moumouvirus battle49	B	0	0
Moumouvirus saoudian	B	0	1
Moumouvirus boug1	B	0	0
Megavirus sete2	C	0	0
Megavirus shan	C	0	0
Megavirus terrain en constructi	C	0	0
Megavirus terra1	C	0	0
Megavirus courdo5	C	0	1
Megavirus courdo7	C	0	0
Megavirus courdo11	C	0	0
Megavirus bus	C	0	1
Megavirus LBA111	C	0	0
Megavirus montpellier3	C	0	0
Megavirus mont1	C	0	0
Megavirus chilensis	C	1	1
Megavirus avenue9	C	0	1
Megavirus balcon	C	0	0
Megavirus potage	C	0	0
Megavirus ursino	C	0	1
Megavirus afrovirus	C	0	0
Megavirus T1	C	0	0
Megavirus battle43	C	0	0
Megavirus T4	C	0	0
Megavirus montadette1	C	0	0
Megavirus feuillage	C	0	0
Megavirus J3	C	0	0
Megavirus T6	C	0	0

Lineage	Number of Mimiviridae genomes	Zamilon inserted sequence (28 bp length)	Zamilon repeat CRISPR-like hits (4 x 15 bp length)
A	28	28/28	28/28
B	8	0/8	0/8
C	24	1/24	0/24

All genomes were screened for the presence of the 28-nucleotide-long stretch (AATCTGATAATGAATCTGATAATGAATC) and the repeated 15-nucleotide sequence (TGATAATGAATCTGA).

Extended Data Table 2 | Functional inferences of the open reading frames in the vicinity of the MIMIVIRE locus in mimivirus

ORFs	Proposed gene-like name	Accession	E-value	Function of ORFs
R343	RNAIII-Like	cd00593	6.85e-31	Probable ribonuclease 3
R350	Helicase domain / SNF2 domain	smart00490 pfam00176	1.66e-04 4.45e-14	Helicase Transcription regulation DNA repair
R354	Cas4-Like	pfam09588	4.69e-34	Putative phage-type endonuclease
L364	Helicase C DEXDc Cas3_I	cd00079 cd00046 cd09639	6.34e-13 2.45e-06 3.97e-03	Putative ATP-dependent RNA helicase

Sequence and structural similarity searches were performed by using BLAST and PHYRE2.

NAFLD causes selective CD4⁺ T lymphocyte loss and promotes hepatocarcinogenesis

Chi Ma¹, Aparna H. Kesarwala², Tobias Eggert¹, José Medina-Echeverz¹, David E. Kleiner³, Ping Jin⁴, David F. Stroncek⁴, Masaki Terabe⁵, Veena Kapoor⁶, Mei ElGindi¹, Miaojun Han¹, Angela M. Thornton⁷, Haibo Zhang⁸, Michèle Egger⁹, Ji Luo⁸, Dean W. Felsher¹⁰, Daniel W. McVicar¹¹, Achim Weber⁹, Mathias Heikenwalder^{12,13} & Tim F. Greten¹

Hepatocellular carcinoma (HCC) is the second most common cause of cancer-related death. Non-alcoholic fatty liver disease (NAFLD) affects a large proportion of the US population and is considered to be a metabolic predisposition to liver cancer^{1–5}. However, the role of adaptive immune responses in NAFLD-promoted HCC is largely unknown. Here we show, in mouse models and human samples, that dysregulation of lipid metabolism in NAFLD causes a selective loss of intrahepatic CD4⁺ but not CD8⁺ T lymphocytes, leading to accelerated hepatocarcinogenesis. We also demonstrate that CD4⁺ T lymphocytes have greater mitochondrial mass than CD8⁺ T lymphocytes and generate higher levels of mitochondrially derived reactive oxygen species (ROS). Disruption of mitochondrial function by linoleic acid, a fatty acid accumulated in NAFLD, causes more oxidative damage than other free fatty acids such as palmitic acid, and mediates selective loss of intrahepatic CD4⁺ T lymphocytes. *In vivo* blockade of ROS reversed NAFLD-induced hepatic CD4⁺ T lymphocyte decrease and delayed NAFLD-promoted HCC. Our results provide an unexpected link between lipid dysregulation and impaired anti-tumour surveillance.

HCC commonly arises in patients with underlying chronic liver disease, and is considered a typical inflammation-associated tumour¹. Recent epidemiology studies indicate an increase in the rate of NAFLD-induced HCC^{2–5}. Immune evasion mediated by numerous immune suppressor mechanisms involving different immune cell subsets have been shown to contribute to HCC initiation and progression⁶, and patients with tumours containing lymphocytic infiltrates show longer survival and lower risk of recurrence⁷. However, the role of adaptive immune responses in NAFLD and HCC have just begun to be understood⁸.

Here we investigated how metabolic changes observed in NAFLD promoted hepatocarcinogenesis using a series of different mouse NAFLD and HCC models and confirmed our results using human samples. Inducible liver-specific MYC oncogene transgenic mice (MYC-ON)⁹ were fed with a methionine-choline-deficient diet (MCD) to induce NAFLD¹⁰ (Fig. 1a and Extended Data Fig. 1a, b). Earlier microscopic liver tumour lesions were found in MYC-ON MCD mice (Fig. 1b, top). As expected, MYC-ON MCD mice showed more macroscopic liver tumours (Fig. 1b, bottom, and c). Similar results were obtained in MYC-ON mice fed a choline-deficient and amino-acid-defined diet (CDAA), another NAFLD-inducing diet (Extended Data Fig. 1c–e)¹¹. Again, more liver tumours were found in diethylnitrosamine carcinogen-challenged C57BL/6 mice^{12,13} fed with a CDAA or high-fat (HF) diet (Extended Data Fig. 1f–i). These results clearly

demonstrate that diet-induced NAFLD enhances HCC in different mouse hepatocarcinogenesis and NAFLD models.

Next, we studied the immune cell subsets in mice with NAFLD and HCC. Consistent with previous reports, dendritic cells, macrophages and CD11b⁺Gr1⁺ cells increased (Extended Data Fig. 2a, b)^{14–16}. Unexpectedly, significantly fewer CD3^{hi}CD4⁺ T lymphocytes, which corresponded to conventional intrahepatic CD4⁺ T lymphocytes, were found in mice with NAFLD (Fig. 1d and Extended Data Fig. 2a–c, e). No significant difference of intrahepatic CD3^{lo}CD4⁺ cells, representing

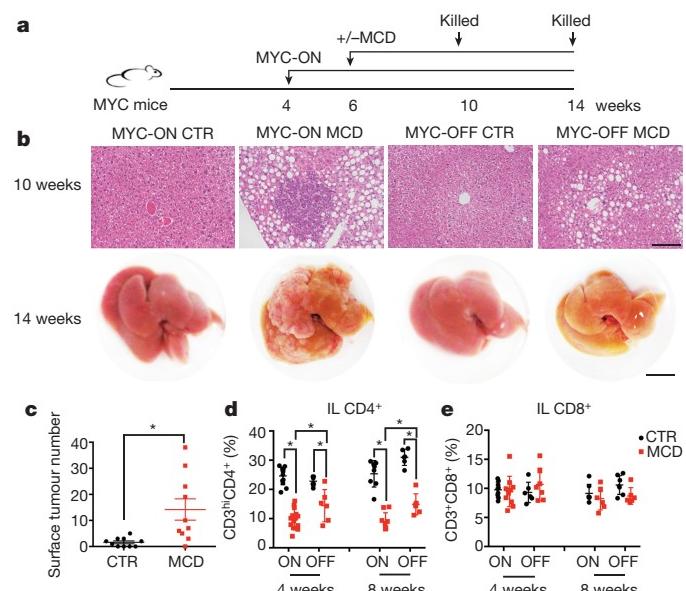


Figure 1 | NAFLD induces a selective loss of intrahepatic CD4⁺ T lymphocytes and promotes HCC. **a**, Experimental setup. **b**, Top, representative haematoxylin and eosin (H&E) liver sections. Scale bar, 10 μm. Bottom, representative liver images. Scale bar, 10 mm. **c**, Liver surface tumour counts. CTR, control diet. $n = 10$ for CTR, 17 for MCD, $P = 0.0067$, Student's *t*-test. **d, e**, Intrahepatic CD4⁺ T lymphocytes (IL CD4⁺) and intrahepatic CD8⁺ T lymphocytes (IL CD8⁺) were measured by flow cytometry. ON, MYC-ON; OFF, MYC-OFF. $n = 12$ for ON-CTR 4 weeks, 15 for ON-MCD 4 weeks, 6 for OFF-CTR 4 weeks, 6 for OFF-MCD 4 weeks, 8 for ON-CTR 8 weeks, 9 for ON-MCD 8 weeks, 6 for OFF-CTR 8 weeks, 6 for OFF-MCD 8 weeks. * $P < 0.05$, two-way analysis of variance (ANOVA). All data are mean ± standard error of the mean (s.e.m.).

¹Gastrointestinal Malignancy Section, Thoracic and Gastrointestinal Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ²Radiation Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ³Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁴Cell Processing Section, Department of Transfusion Medicine, Clinical Center, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁵Vaccine Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁶Experimental Transplantation and Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁷Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁸Laboratory of Cancer Biology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁹Institute of Surgical Pathology, University and University Hospital Zurich, Zurich 8091, Switzerland. ¹⁰Division of Oncology, Department of Medicine and Pathology, Stanford University, California 94305, USA. ¹¹Cancer and Inflammation Program, National Cancer Institute, Frederick, Maryland 21702, USA. ¹²Institute of Virology, Technische Universität München/Helmholtz Zentrum München, Munich 81675, Germany. ¹³Division of Chronic Inflammation and Cancer, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany.

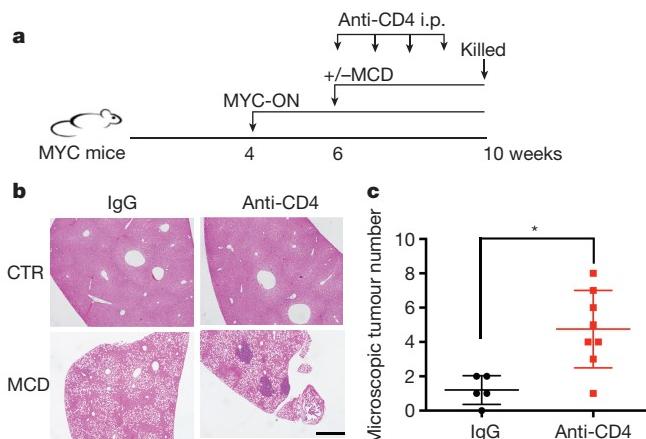


Figure 2 | Depletion of intrahepatic CD4⁺ T lymphocytes accelerates tumour development in MYC-ON MCD mice. **a**, Experimental setup. i.p., intraperitoneal. **b**, **c**, Representative H&E staining images and microscopic tumour counts. Scale bar, 200 μ m. Data are mean \pm s.e.m.; $n = 5$ for IgG, 8 for anti-CD4. * $P < 0.05$, Student's *t*-test.

natural killer (NK) T cells (Extended Data Fig. 2d, f), or splenic CD4⁺ T lymphocytes was observed (Extended Data Fig. 2g). Unlike CD4⁺ T lymphocytes, intrahepatic CD8⁺ T lymphocytes remained unchanged (Fig. 1e and Extended Data Fig. 2a, b). The liver-specific reduction of CD4⁺ but not CD8⁺ T lymphocytes was also observed in the two other dietary NAFLD models in both tumour-free and tumour-bearing settings (Extended Data Fig. 2h–t), illustrating a tumour-independent but NAFLD-dependent mechanism. In addition, fewer CD4⁺ but not CD8⁺ T lymphocytes were found in leptin-deficient (*ob/ob*) mice (Extended Data Fig. 2u–x).

CD4⁺ T lymphocytes in NAFLD mice were characterized. Higher levels of CD69 and CD44^{hi}CD62^{lo} subsets were found in MYC-ON MCD mice (Extended Data Fig. 3a–d). Hepatic but not splenic CD4⁺ T lymphocytes also consistently produced more interferon (IFN)- γ but not interleukin (IL)-4 (Extended Data Fig. 3e–g). Although T-bet, GATA3 and Foxp3 frequency did not change, more ROR- γ t was detected in MYC-ON MCD mice (Extended Data Fig. 3h). Accordingly, intrahepatic CD4⁺ T lymphocytes produced more IL-17 in NAFLD (Extended Data Fig. 3i, j). No change of regulatory T lymphocyte (T_{reg}) frequency was found, and absolute numbers decreased (Extended Data Fig. 3h, k), consistent with a previous report¹⁷. In addition, the T_{reg} function remained unchanged in NAFLD (Extended Data Fig. 3l). Together, our results indicate that NAFLD caused activation of hepatic CD4⁺ T lymphocytes in mice.

CD4⁺ T lymphocytes have been reported to inhibit HCC initiation and mediate tumour regression^{18,19}. It has also been reported that a considerable fraction of non-synonymous cancer mutations is immunogenic and that the majority of the immunogenic mutanome is recognized by CD4⁺ T lymphocytes²⁰. Therefore, we studied tumour-specific CD4⁺ T lymphocytes. α -Fetoprotein (AFP)-specific CD4⁺ T lymphocytes in MYC-ON MCD mice were detected, suggesting that MYC tumours induced anti-tumour CD4⁺ T lymphocyte responses (Extended Data Fig. 3m). Next, we depleted CD4⁺ T lymphocytes to study their relevance to tumour growth. CD4 antibody depletion (Extended Data Fig. 3n) caused more hepatic tumour lesions in MYC-ON MCD mice (Fig. 2a–c). In control-diet-fed MYC mice, CD4 antibody depletion also promoted tumours but at a later time point (Extended Data Fig. 3o, p). These results suggested that loss of CD4⁺ T lymphocytes strongly contributed to HCC development in MYC-ON mice.

Next, we studied CD4⁺ T lymphocyte survival in mice with NAFLD, and a higher annexin V⁺ level was found in MYC-ON MCD mice than that in MYC-ON mice fed with control diet (Fig. 3a and Extended Data Fig. 4a). We hypothesized that the extensive hepatic lipid accumulation

induced the CD4⁺ T lymphocyte loss. Hepatocytes isolated from MYC-ON MCD mice, which showed accumulation of lipid droplets (Extended Data Fig. 4b), or control mice, were co-cultured with splenocytes. Interestingly, a significant increase in annexin V⁺7AAD⁺ cells was seen in CD4⁺ but not CD8⁺ T lymphocytes (Extended Data Fig. 4c–e). No cell-to-cell contact was required (Fig. 3b). Higher lipid levels were detected in hepatic CD4⁺ T lymphocytes in MCD mice (Extended Data Fig. 4f, g). This prompted us to examine whether lipids released from lipid-laden hepatocytes were taken up by CD4⁺ T lymphocytes and caused cell death. To test this hypothesis, we first measured the hepatic free fatty acid (FFA) composition. Consistent with previous reports, palmitic acid (C16:0), stearic acid (C18:0), linoleic acid (C18:2), arachidonic acid (C20:4) and docosahexaenoic acid (C22:6) are the abundant FFAs (Fig. 3c and Extended Data Table 1). Although the total amount of FFAs did not change significantly (Extended Data Table 1), the levels of C16:0 and C22:6 decreased. C18:2 was the only abundant FFA, which accumulated in the liver after MCD treatment (Fig. 3c). Our data are supported by previous reports of hepatic C18:2 accumulation in HF-diet-induced NAFLD mice and *ob/ob* mice^{8,21}. Next, we depleted FFAs from conditioned hepatocyte-culture medium. As expected, FFA-depleted conditioned medium no longer caused CD4⁺ T lymphocyte death (Fig. 3d). FFAs in conditioned medium from lipid-laden hepatocytes were further analysed, and C16:0, C18:0 and C18:2 were identified as predominant FFAs (Extended Data Fig. 4h, i).

Then, isolated CD4⁺ T lymphocytes were incubated with individual FFAs to study their effect on cell survival. Comparing other FFAs, C18:2 treatment caused a substantially higher level of 7AAD⁺annexin V⁺ cells (Fig. 3e). Unlike CD4⁺ T lymphocytes, cell death in CD8⁺ T lymphocytes was not affected at the tested concentration (Fig. 3f). Similar results were found in activated T lymphocytes (Extended Data Fig. 4j). Dose-response and time-course analysis confirmed that CD4⁺ T lymphocytes were more susceptible to C18:2-induced cell death than CD8⁺ T lymphocytes (Extended Data Fig. 4k–m). The increase of caspase 3/7 activity confirmed that CD4⁺ T lymphocytes died through apoptosis (Extended Data Fig. 4n). Similar cell death rates between CD4⁺ and CD8⁺ T lymphocytes were observed in an H₂O₂-induced cell death model, showing that the effect is specific to C18:2 (Extended Data Fig. 4o). Interestingly, mice fed with a high C18:2 diet showed a reduction in CD4⁺ but not CD8⁺ T lymphocytes (Fig. 3g, h) suggesting that C18:2 is sufficient to cause CD4⁺ T lymphocyte death *in vivo*.

The mechanism of how C18:2 induced CD4⁺ T lymphocyte death was studied. No difference in cellular C18:2 uptake by CD4⁺ versus CD8⁺ T lymphocytes was found (Extended Data Fig. 4p). Direct assessment revealed greater mitochondrial mass in CD4⁺ lymphocytes (Fig. 4a). Microarray analysis revealed that oxidative phosphorylation and mitochondrial dysfunction pathways were specifically altered in CD4⁺ but not CD8⁺ T lymphocytes after C18:2 treatment (Extended Data Fig. 5). CPT1a, the rate-limiting enzyme for importing FFAs into mitochondria, increased in parallel with the decrease of a number of genes coding components for the electron transport complex (Extended Data Fig. 6a). C18:2 was more potent than other FFAs in upregulating CPT1a (Fig. 4b). A similar effect was observed in Jurkat cells, a human CD4⁺-derived T leukaemia cell line (Extended Data Fig. 6b). Knockdown of CPT1a rescued Jurkat cells from C18:2-induced cell death (Fig. 4c and Extended Data Fig. 6c). All these results pointed towards mitochondria as the critical mediator for CD4⁺ T lymphocyte death. Inside mitochondria, FFAs are β -oxidized to fuel ATP generation via the electron transport chain (ETC). Fatty acid oxidation (FAO) was measured, and C18:2 showed a greater FAO rate than C16:0 (Fig. 4d). Higher FAO favours more NADH entering the ETC to generate ATP. However, our array data also suggested that C18:2 impaired ETC function (Extended Data Fig. 6a). Indeed, mitochondrial membrane potential, which is maintained by proper ETC activity, was significantly decreased by C18:2 in CD4⁺ but not CD8⁺ T lymphocytes (Fig. 4e). A disrupted ETC can become a major site

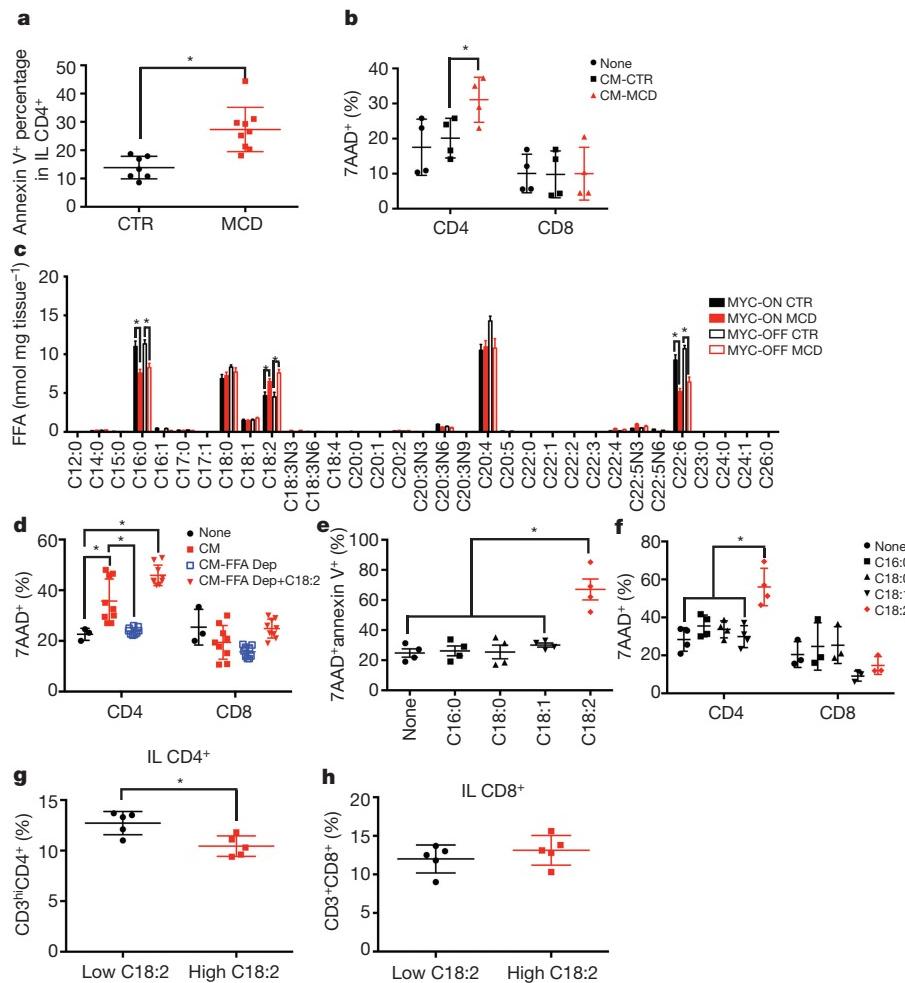


Figure 3 | Lipid-laden hepatocytes cause CD4⁺ T lymphocyte death through releasing C18:2. **a**, *Ex vivo* cell death of intrahepatic CD4⁺ T lymphocytes from MYC-ON NAFLD mice ($n=7$ for CTR, $n=9$ for MCD, Student's *t*-test). **b**, Lymphocyte survival after incubation with hepatocyte conditioned medium (CM) ($n=4$, two-way ANOVA). **c**, Hepatic total FFA composition analysis ($n=6$, * $P<0.05$, ANOVA). **d**, FFA depletion (Dep)

of premature electron leakage to oxygen to generate ROS and lead to cell death²².

To assess mitochondrial respiration, oxygen consumption analysis was performed. Normalized oxygen consumption rates (OCRs) were significantly higher in CD4⁺ T lymphocytes compared with CD8⁺ T lymphocytes, consistent with previous reports²³. Treatment with oligomycin, an inhibitor of mitochondrial ATP synthase, revealed substantial levels of ATP-synthase-dependent oxygen consumption in both CD4⁺ and CD8⁺ T lymphocytes (Fig. 4f and Extended Data Fig. 6d). C18:2 abrogated the oligomycin-sensitive fraction of the OCR in CD4⁺ and CD8⁺ T lymphocytes without reducing total oxygen consumption levels (Fig. 4f and Extended Data Fig. 6d). These data are consistent with a shift in oxygen consumption from ATP-synthase-dependent to ATP-synthase-independent ROS production. In contrast, C16:0 failed to eliminate ATP-synthase-dependent oxygen consumption in CD4⁺ T lymphocytes (Fig. 4g and Extended Data Fig. 6e). Consistently, increased total ROS production was found in CD4⁺ T lymphocytes when co-cultured with C18:2 versus C16:0 (Extended Data Fig. 6f). Moreover, elevated ROS levels were detected in hepatic CD4⁺ T lymphocytes *ex vivo* under NAFLD conditions (Fig. 4h). Finally, mitochondrial superoxide was confirmed to increase selectively in CD4⁺ T lymphocytes after C18:2 treatment (Fig. 4i), and CPT1a knockdown blocked C18:2-induced mitochondrial ROS production (Extended Data Fig. 6g). Taken together, these data

from conditioned medium ($n=3$ for none, $n=9$ for other treatments; two-way ANOVA). **e**, **f**, Lymphocyte survival after FFA treatment ($n=4$, one-way ANOVA). **g**, **h**, CD4⁺ and CD8⁺ T lymphocytes in high- or low-C18:2-diet-fed mice ($n=5$, Student's *t*-test). All data are mean \pm s.e.m. * $P<0.05$.

suggest that greater levels of mitochondrial-derived ROS accumulate in CD4⁺ T lymphocytes after C18:2 treatment, leading to their depletion.

Therefore, we tested the role of ROS in NAFLD-associated CD4⁺ T lymphocyte death and HCC development *in vivo*. Blocking ROS with catalase or N-acetylcysteine (NAC) abrogated cell death *in vitro* in CD4⁺ T lymphocytes when incubated with hepatocytes from MCD-ON MYC mice (Fig. 4j). Similarly, catalase and NAC prevented C18:2-induced CD4⁺ T lymphocyte death *in vitro* (Extended Data Fig. 6h). Oxidative stress is an important factor in NAFLD progression²⁴. To test whether ROS mediate hepatic CD4⁺ T lymphocyte loss *in vivo*, we treated MCD-diet-fed mice with NAC. Although NAC treatment did not influence steatosis (Extended Data Fig. 6i, j), it effectively reversed the loss of hepatic CD4⁺ T lymphocytes (Fig. 4k). More importantly, NAC treatment significantly delayed NAFLD-promoted tumour development (Fig. 4l and Extended Data Fig. 6j). Tumour lesions occurred despite NAC treatment when CD4⁺ T lymphocytes were removed, suggesting that prevention of CD4⁺ T lymphocyte death mediates at least partially the anti-tumour effect of NAC. Similar results were obtained using mitoTEMPO²⁵, a specific mitochondrial antioxidant in both *in vitro* and *in vivo* settings (Fig. 4m-o).

C18:2 has also been identified as an important fatty acid in the context of NAFLD in humans^{26,27}. We tested whether C18:2 also affects human CD4⁺ T lymphocyte survival. Consistent with our mouse data,

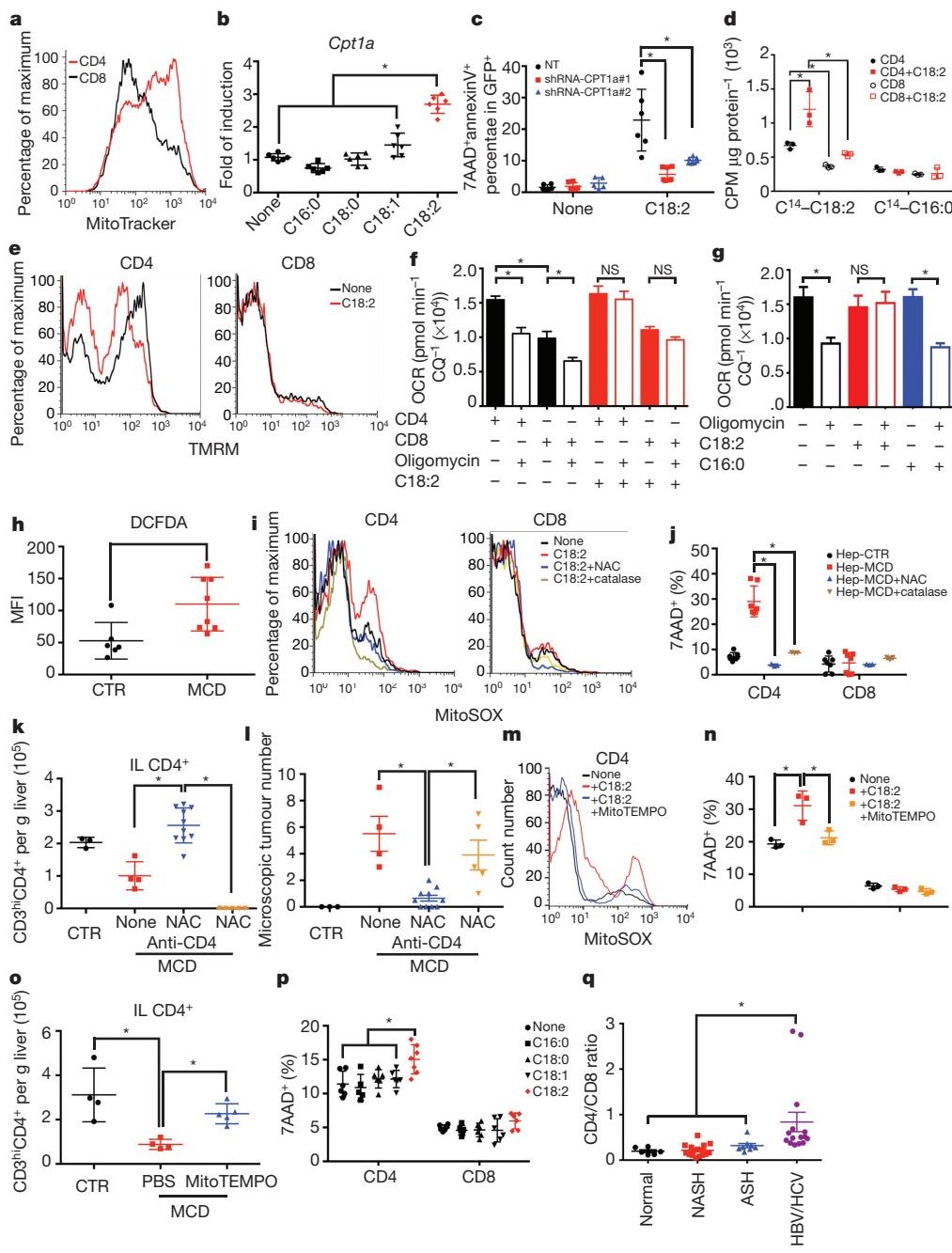


Figure 4 | Mitochondrial ROS mediates C18:2-induced CD4⁺ T lymphocyte death. **a**, Mitochondrial mass analysis. **b**, *Cpt1a* mRNA levels in FFA-treated CD4⁺ T lymphocytes ($n=6$). **c**, CPT1a knockdown on C18:2-induced Jurkat cell death ($n=6$). shRNA, short hairpin RNA. NT, non-targeting control. **d**, Oxidation rate of C18:2 or C16:0 in lymphocytes ($n=3$). CPM, counts per minutes. **e**, Mitochondrial membrane potential in C18:2-treated lymphocytes. TMRM, tetramethylrhodamine, methyl ester. **f**, **g**, OCR assay of activated CD4⁺ and CD8⁺ T lymphocytes treated with FFAs ($n=8$). CQ, arbitrary unit obtained by CyQUANT cell proliferation assay (see Methods). **h**, *Ex vivo* ROS levels of intrahepatic CD4⁺ T lymphocytes ($n=6$ for CTR, 8 for MCD). DCFDA, 2',7'-dichlorofluorescin diacetate. MFI, mean fluorescence intensity.

C18:2, but no other tested FFAs, caused selective CD4⁺ but not CD8⁺ T lymphocyte death (Fig. 4p and Extended Data Fig. 7a). Similarly, C18:2 but not C16:0 increased the ROS level in human CD4⁺ T lymphocytes (Extended Data Fig. 7b). Finally, intrahepatic CD4⁺ T lymphocytes in liver biopsies from patients with non-alcoholic steatohepatitis (NASH), alcoholic steatohepatitis (ASH) and viral hepatitis

were determined (Extended Data Table 2). While ALT and AST levels did not differ among patients with different liver diseases (Extended Data Fig. 7c, d), fewer CD4⁺ T lymphocytes were found in NASH and ASH patients than in viral hepatitis patients (Extended Data Figs 7e and 8), and the CD4/CD8 ratio was significantly lower in NASH patients, supporting the idea of selective CD4⁺ T lymphocyte loss (Fig. 4q).

Interestingly, lower CD4 counts were also found in ASH patients; ASH has very similar histological features to NASH.

Dysregulation of lipid metabolism and accumulation of lipids in the liver is part of the aetiology of NAFLD. So far, NASH has been described as causing NF- κ B dysregulation, activation of the inflammasome, Toll-like receptor activation and affecting innate immune responses through multiple pathways or directly affecting hepatocytes^{28–30}. Our results extend these findings by describing a novel link between obesity-induced lipid accumulation and selective CD4 $^{+}$ T lymphocyte loss, suggest a critical role for CD4 $^{+}$ T lymphocytes in the disease progression from NAFLD to HCC.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 November 2015; accepted 4 January 2016.

Published online 2 March 2016.

1. European Association For The Study Of The Liver & European Organisation For Research And Treatment Of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J. Hepatol.* **56**, 908–943 (2012).
2. Sun, B. & Karin, M. Obesity, inflammation, and liver cancer. *J. Hepatol.* **56**, 704–713 (2012).
3. Michelotti, G. A., Machado, M. V. & Diehl, A. M. NAFLD, NASH and liver cancer. *Nature Rev. Gastroenterol. Hepatol.* **10**, 656–665 (2013).
4. Schuppan, D. & Schattenberg, J. M. Non-alcoholic steatohepatitis: pathogenesis and novel therapeutic approaches. *J. Gastroenterol. Hepatol.* **28** (suppl. 1), 68–76 (2013).
5. Wree, A., Broderick, L., Canbay, A., Hoffman, H. M. & Feldstein, A. E. From NAFLD to NASH to cirrhosis—new insights into disease mechanisms. *Nature Rev. Gastroenterol. Hepatol.* **10**, 627–636 (2013).
6. Greten, T. F., Wang, X. W. & Korangy, F. Current concepts of immune based treatments for patients with HCC: from basic science to novel treatment approaches. *Gut* **64**, 842–848 (2015).
7. Greten, T. F., Duffy, A. G. & Korangy, F. Hepatocellular carcinoma from an immunologic perspective. *Clin. Cancer Res.* **19**, 6678–6685 (2013).
8. Wolf, M. J. et al. Metabolic activation of intrahepatic CD8 $^{+}$ T cells and NKT cells causes nonalcoholic steatohepatitis and liver cancer via cross-talk with hepatocytes. *Cancer Cell* **26**, 549–564 (2014).
9. Shachaf, C. M. et al. MYC inactivation uncovers pluripotent differentiation and tumour dormancy in hepatocellular cancer. *Nature* **431**, 1112–1117 (2004).
10. Rinella, M. E. et al. Mechanisms of hepatic steatosis in mice fed a lipogenic methionine choline-deficient diet. *J. Lipid Res.* **49**, 1068–1076 (2008).
11. Yang, L. et al. Transforming growth factor β signalling in hepatocytes participates in steatohepatitis through regulation of cell death and lipid metabolism in mice. *Hepatology* **59**, 483–495 (2014).
12. Park, E. J. et al. Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. *Cell* **140**, 197–208 (2010).
13. Kapanadze, T. et al. Regulation of accumulation and function of myeloid derived suppressor cells in different murine models of hepatocellular carcinoma. *J. Hepatol.* **59**, 1007–1013 (2013).
14. Xia, S. et al. Gr-1 $^{+}$ CD11b $^{+}$ myeloid-derived suppressor cells suppress inflammation and promote insulin sensitivity in obesity. *J. Biol. Chem.* **286**, 23591–23599 (2011).
15. Baeck, C. et al. Pharmacological inhibition of the chemokine CCL2 (MCP-1) diminishes liver macrophage infiltration and steatohepatitis in chronic hepatic injury. *Gut* **61**, 416–426 (2012).
16. Henning, J. R. et al. Dendritic cells limit fibroinflammatory injury in nonalcoholic steatohepatitis in mice. *Hepatology* **58**, 589–602 (2013).
17. Ma, X. et al. A high-fat diet and regulatory T cells influence susceptibility to endotoxin-induced liver injury. *Hepatology* **46**, 1519–1529 (2007).
18. Rakha, K. et al. CD4 $^{+}$ T cells contribute to the remodeling of the microenvironment required for sustained tumor regression upon oncogene inactivation. *Cancer Cell* **18**, 485–498 (2010).
19. Kang, T. W. et al. Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature* **479**, 547–551 (2011).
20. Kreiter, S. et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* **520**, 692–696 (2015).
21. Giesbertz, P. et al. Metabolite profiling in plasma and tissues of *ob/ob* and *db/db* mice identifies novel markers of obesity and type 2 diabetes. *Diabetologia* **58**, 2133–2143 (2015).
22. Schönfeld, P. & Wojtczak, L. Fatty acids as modulators of the cellular production of reactive oxygen species. *Free Radic. Biol. Med.* **45**, 231–241 (2008).
23. Cao, Y., Rathmell, J. C. & Macintyre, A. N. Metabolic reprogramming towards aerobic glycolysis correlates with greater proliferative ability and resistance to metabolic inhibition in CD8 versus CD4 T cells. *PLoS ONE* **9**, e104104 (2014).
24. Sumida, Y., Niki, E., Naito, Y. & Yoshikawa, T. Involvement of free radicals and oxidative stress in NAFLD/NASH. *Free Radic. Res.* **47**, 869–880 (2013).
25. Porporato, P. E. et al. A mitochondrial switch promotes tumor metastasis. *Cell Reports* **8**, 754–766 (2014).
26. Feldstein, A. E. et al. Mass spectrometric profiling of oxidized lipid products in human nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *J. Lipid Res.* **51**, 3046–3054 (2010).
27. Muir, K. et al. Proteomic and lipidomic signatures of lipid metabolism in NASH-associated hepatocellular carcinoma. *Cancer Res.* **73**, 4722–4731 (2013).
28. Beraza, N. et al. Pharmacological IKK2 inhibition blocks liver steatosis and initiation of non-alcoholic steatohepatitis. *Gut* **57**, 655–663 (2008).
29. Henao-Mejia, J. et al. Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* **482**, 179–185 (2012).
30. Miura, K. et al. Toll-like receptor 2 and palmitic acid cooperatively contribute to the development of nonalcoholic steatohepatitis through inflammasome activation in mice. *Hepatology* **57**, 577–589 (2013).

Acknowledgements We would like to thank J. Berzofsky, W. Stoffel, E. M. Shevach and S. Thorgeirsson for helpful discussion. A.M.T. was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases. J.L. was supported by the Intramural Program Grant ZIAAC011303, NIH, National Cancer Institute (NCI). M.H. was supported by a European Research Council starting grant (LiverCancerMechanism), the Stiftung Experimentelle Biomedizin (Hofsneider Stiftung), the Pre-clinical Comprehensive Center (PCCC) and the Helmholtz foundation. A.W. was supported by grants from the Krebsliga Schweiz (Oncosuisse) and the Promedica Stiftung, Switzerland. D.E.K., A.H.K., D.W.M. and T.F.G. were supported by the Intramural Research Program of the NIH, NCI.

Author Contributions C.M. and A.H.K. performed experiments. C.M., A.H.K., D.W.M. and T.F.G. analysed data. T.E., J.M.-E., D.E.K., P.J., D.F.S., M.T., V.K., M.E., M.H., A.M.T., H.Z., J.L. and D.F.W. assisted with experiments and analysis or provided valuable reagents. A. H.K. and D.W.M. performed and analysed seahorse studies. M.E., A.W. and M.H. collected/provided human specimens and performed/analysed immunohistochemistries of human samples. C.M. and T.F.G. conceived of and designed the project. C.M. and T.F.G. wrote the manuscript and all authors contributed to writing and provided feedback.

Author Information Microarray data have been deposited in the Gene Expression Omnibus under accession number GSE67918. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.F.G. (tim.greten@nih.gov).

METHODS

Mouse studies. LAP-tTA and TRE-MYC mice were previously described and MYC expression in the liver was activated by removing doxycycline treatment ($100\text{ }\mu\text{g ml}^{-1}$) from the drinking water of 4-week-old double transgenic mice for both TRE-MYC and LAP-tTA as previously described^{9,13}. C57BL/6 mice were obtained from NCI Frederick. Chemically induced HCC was established by intra-peritoneal injection of diethylnitrosoamine (DEN) (Sigma) into 2-week-old male pups at a dose of $20\text{ }\mu\text{g g}^{-1}$ body weight¹³. Twelve-week-old male B6.Cg-Lep^{ob}/J (*ob/ob*) mice or wild-type control mice were obtained from Charles River. Foxp3-GFP mice were previously described³¹. NAFLD was induced by feeding mice with a methionine-choline-deficient (MCD) diet (catalogue number 960439, MP biomedical), a choline-deficient and amino-acid-defined (CDA) diet (catalogue number 518753, Dyets) or a high-fat diet (catalogue number F3282, Bio Serv) for the indicated time^{10,11,32}. The MCD diet was supplied with corn oil (10%, w/w), and no fish oil was added. Control diet was purchased from MP Biomedical (catalogue number 960441). Custom-made high- or low-linoleic-acid mouse diets were purchased from Research Diets. The modified diets were based on AIN-76A standard mouse diet, and are isocaloric (4.45 kcal g^{-1}) and contained the same high-fat content (23%, w/w). Linoleic-acid-rich safflower oil and saturated fatty-acid-containing coconut oil were supplied at different ratios to yield 2% (w/w) for the low-linoleic-acid diet or 12% (w/w) for the high-linoleic-acid diet. C57BL/6 mice were fed with the high- or low-linoleic-acid diet for 4 weeks. MYC mice were injected i.p. with $50\text{ }\mu\text{g}$ CD4 antibody (clone GK1.5, BioXcell) every week for the indicated time period to deplete CD4⁺ T cells³³. N-acetylcysteine (NAC) was given in drinking water (10 mg ml^{-1})³⁴ for the indicated time period to prevent excess ROS production. Mitochondrial-specific antioxidant mitoTEMPO was purchased from Sigma. Mice received mitoTEMPO at a dose of 0.7 mg kg^{-1} per day²⁵ by osmotic minipumps (ALZET). At the experimental end points, mice were killed. For flow cytometry analysis, single-cell suspensions were prepared from spleen, liver and blood as described previously. Red blood cells were lysed by ACK Lysis Buffer (Quality Biologicals). Parts of live tissue were fixed by 10% formaldehyde and subjected to H&E staining. Free fatty acids were purchased from Sigma.

Oil Red O staining. Lipid accumulation was detected by Oil Red O staining in frozen liver sections using the custom service of Histo Serv.

Flow cytometry. Cells were surface-labelled with the indicated antibodies for 15 min at 4°C . Flow cytometry was performed on BD FACSCalibur or BD LSRII platforms and results were analysed using FlowJo software version 9.3.1.2 (TreeStar). The following antibodies were used for flow cytometry analysis: anti-CD3-FITC (clone 17A2, BD Pharmingen), anti-CD4-PE (clone RM4-4, Biolegend), anti-CD4-APC (clone RM4-5, eBioscience), anti-CD8-Alexa Fluor 700 (clone 53-6.7 Biolegend), anti-CD45, anti-CD44-PE (clone IM7, eBioscience), anti-CD62L-PerCP/Cy5.5 (MEL-14, Biolegend), anti-CD69-Pacific blue (clone H1.2F3, Biolegend), PBS57/CD1d-tetramer-APC (NIH core facility). To determine cytokine production, cells were stimulated with PMA and ionomycin for 30 min, and then were fixed and permeabilized using cytofix/cytoperm kit (BD Pharmingen) followed by anti-IFN- γ -PE (clone XMG1.2, BD Pharmingen), anti-IL-17-PerCP/Cy5.5 (clone TC11-18H10.1, Biolegend) staining. Cell death and apoptosis were detected with annexin V-PE (BD Pharmingen) and 7-AAD (BD Pharmingen) staining according to the manufacturer's instructions. Intrahepatic CD4⁺ lymphocytes were gated on the CD3^{hi}CD4⁺ population from total live hepatic infiltrating mononuclear cells. Absolute numbers were calculated by multiplying frequencies obtained from flow by total live mononuclear cell count, then divided by liver weight. The antibodies used for human peripheral blood mononuclear cell (PBMC) staining are the following: anti-CD3-PE (clone SK7, BD Pharmingen), anti-CD4-FITC (clone RPA-T4, BD Pharmingen), anti-CD8-APC (clone RPA-T8, BD Pharmingen).

T_{reg} suppressive function assay. Murine T_{reg} assays were performed as described³¹. Briefly, liver T_{reg} cells were isolated as CD4⁺GFP⁺ by flow-cytometry-assisted cell sorting from Foxp3-GFP mice kept on an MCD or control diet for 4 weeks. CD4⁺GFP⁺ T effector (T_{eff}) cells (5×10^4) were stimulated for 72 h in the presence of irradiated T-depleted splenocytes (5×10^4) plus CD3ε monoclonal antibody ($1\text{ }\mu\text{g ml}^{-1}$), with or without T_{reg} cells added at different ratios. ^3H -Thymidine was added to the culture for the last 6 h and incorporated radioactivity was measured.

AFP-specific T-cell response. Freshly isolated splenocytes from MYC-ON MCD mice were incubated with $5\text{ }\mu\text{g ml}^{-1}$ of mouse α-fetoprotein protein (MyBioSource) for 24 h. GolgiPlug was added for the last 6 h. Then, cells were fixed and permeabilized using cytofix/cytoperm kit (BD Pharmingen) followed by anti-IFN- γ -PE (clone XMG1.2, BD Pharmingen) staining.

Hepatocyte isolation. Primary mouse hepatocytes were isolated from MYC mice and cultured according to a previous report³⁵. Briefly, mice were anaesthetized and the portal vein was cannulated under aseptic conditions. The livers were perfused

with EGTA solution (5.4 mM KCl, 0.44 mM KH₂PO₄, 140 mM NaCl, 0.34 mM Na₂HPO₄, 0.5 mM EGTA, 25 mM Tricine, pH 7.2) and Gey's balanced salt solution (Sigma), and digested with 0.075% collagenase solution. The isolated mouse hepatocytes were then cultured with complete RPMI media in collagen-I-coated plates.

Hepatic fatty acid profiling. Hepatic fatty acid composition was measured at LIPID MAPS lipidomics core at the University of California (San Diego) using an esterified and non-esterified (total) fatty acid panel. Briefly, liver tissues were homogenized and lipid fraction was extracted using a modified Bligh Dyer liquid/liquid extraction method. The lipids were saponified and the hydrolysed fatty acids were extracted using a liquid/liquid method. The extracted fatty acids were derivatized using pentaflourylbenzylbromine (PFB) and analysed by gas chromatography (GC) using an Agilent GC/mass spectrometry (MS) ChemStation. Individual analytes were monitored using selective ion monitoring (SIM). Analytes were monitored by peak area and quantified using the isotope dilution method using a deuterated internal standard and a standard curve.

Free fatty acid identification. Isolated primary hepatocytes from MYC mice fed with MCD or control diet were cultured in complete RPMI for 24 h. Supernatant were harvested and FFAs were identified by GC/MS.

Microarray analysis. Splenocytes from MYC mice were cultured with or without $50\text{ }\mu\text{M}$ C18:2 for 24 h. CD4⁺ and CD8⁺ T lymphocytes were sorted and total RNA was extracted using miRNeasy mini kit (Qiagen). Array analysis was performed in the Department of Transfusion Medicine, clinical centre at NIH. Mouse gene 2.0 ST array (Affymetrix) was used and performed according to the manufacturer's instruction. Data were log-transformed (base 2) for subsequent statistical analysis. The Partek Genomic Suite 6.4 was used for the identification of differentially expressed transcripts. The Ingenuity Pathway Analysis tool (<http://www.ingenuity.com>) was used for analysis of functional pathways.

RNA isolation and real-time PCR. RNA was extracted from frozen tissues with RNeasy Mini Kit (Qiagen). Complementary DNA was synthesized by iScript cDNA synthesis kit (BioRad). Sequence of primers used for quantitative RT-PCR can be obtained from the authors. The reactions were run in triplicates using iQSYBR green supermix kit (BioRad). The results were normalized to endogenous GAPDH expression levels.

CD4⁺ T-cell isolation and co-culture with fatty acids. CD4⁺ T lymphocytes were isolated from the spleen of MYC mice by negative autoMACS selection using a CD4⁺ T lymphocytes isolation kit (Miltenyi Biotec) or flow cytometry cell sorting. Human CD4⁺ T lymphocytes were prepared from PBMCs by autoMACS using a CD4⁺ T lymphocytes isolation kit (Miltenyi Biotec). The purity of CD4⁺ T lymphocytes was above 90% after autoMACS separation and above 95% after flow cytometry cell sorting. C16:0, C18:0, C18:1, and C18:2 were purchased from Sigma. Fatty acids were dissolved in DMEM with 2% fatty-acid-free bovine serum albumin (BSA; Sigma, catalogue number A8806) after solvent was evaporated, then followed by two rounds of vortexing and 30 s of sonication. Isolated CD4⁺ T lymphocytes or splenocytes were incubated with different fatty acids or conditioned medium from hepatocyte culture for 3 days. Unless specifically described, fatty acids were used at $50\text{ }\mu\text{M}$ concentration. For fatty acid depletion, active charcoal (catalogue number C-170, Fisher) was used as described before³⁶. Briefly, 0.5 g of active charcoal was added into every 10 ml of conditioned medium. Then pH was lowered to 3.0 by addition of 0.2 N HCl. The solution was rotated at 4°C for 2 h. Charcoal was then removed by centrifugation, and the clarified solution was brought back to pH 7.0 by addition of 0.2 N NaOH. NAC (10 mM), catalase (1,000 U ml⁻¹) or mitoTEMPO (10 μM) was used to inhibit ROS production, mitochondrial ROS levels were determined by mitoSOX staining 24 h after treatment, cell death and apoptosis were measured by annexin V and 7-AAD staining 3 days after treatment.

Caspase activity assay. Caspase activity assay was measured by caspase-Glo 3/7 assay kit (Promega) according to the manufacturer's protocol.

BODIPY staining. Fresh prepared liver-infiltrating mononuclear cells were washed and resuspended in 500 μl of BODIPY 493/503 at $0.5\text{ }\mu\text{g ml}^{-1}$ in PBS. Cells were stained for 15 min at room temperature. Then cells were subjected to flow cytometry analysis.

RNA interference assay. Two pZIP lentiviral shRNA vectors targeting human CPT1a and a control vector (NT#4) were purchased from TransOMIC Technologies. Lentivirus was packed in 293T cells. Jurkat cells were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ), and no authentication test was performed by us. Cells were cultured in complete RPMI medium and were tested to be mycoplasma free. Jurkat cells were infected with shRNA lentivirus. Puromycin was added to eliminate non-transduced cells. Doxycycline (100 ng ml^{-1}) was added to induce shRNA and GFP expression for 3 days. Efficiency of shRNAs was confirmed by western blot. Jurkat cells with CPT1a knockdown were treated with $200\text{ }\mu\text{M}$ C18:2 for 24 h. Mitochondrial ROS production and cell survival were measured in GFP⁺-transduced cells.

Fatty acid oxidation assay. Fatty acid oxidation was measured according to a previous publication³⁷. 1-¹⁴C-C18:2 and 1-¹⁴C-C16:0 were purchased from PerkinElmer. Briefly, isolated CD4⁺ or CD8⁺ T lymphocytes were pretreated with C18:2 or kept in regular media. After 24 h, cell media was changed to media containing 50 μM cold C18:2 plus 1 μCi 1-¹⁴C-C18:2 per ml or 50 μM cold C16:0 plus 1 μCi 1-¹⁴C-C16:0 per ml. After 2 h, medium was removed and mixed with concentrated perchloric acid (final concentration 0.3 M) plus BSA (final concentration 2%) to precipitate the radiolabelled fatty acids. Samples were vortexed and centrifuged (10,000g for 10 min). Radioactivity was determined in the supernatant to measure water-soluble β-oxidation products.

Mitochondrial membrane potential and ROS staining. Mitochondrial membrane potential was measured by TMRM (ImmunoChemistry Technologies) staining according to the manufacturer's protocol. Briefly, cells were kept in culture medium with 100 nM of TMRM for 20 min in a CO₂ incubator at 37 °C. After washing twice, cells were processed to flow cytometry analysis. Mitochondria-associated superoxide was detected by mitoSOX (Life Technologies) staining according to the manufacturer's protocol. Briefly, cells were first subjected to surface marker staining. Then cells were stained with 2.5 μM mitoSOX for 30 min in a CO₂ incubator at 37 °C. After washing twice, cells were processed for flow cytometry analysis.

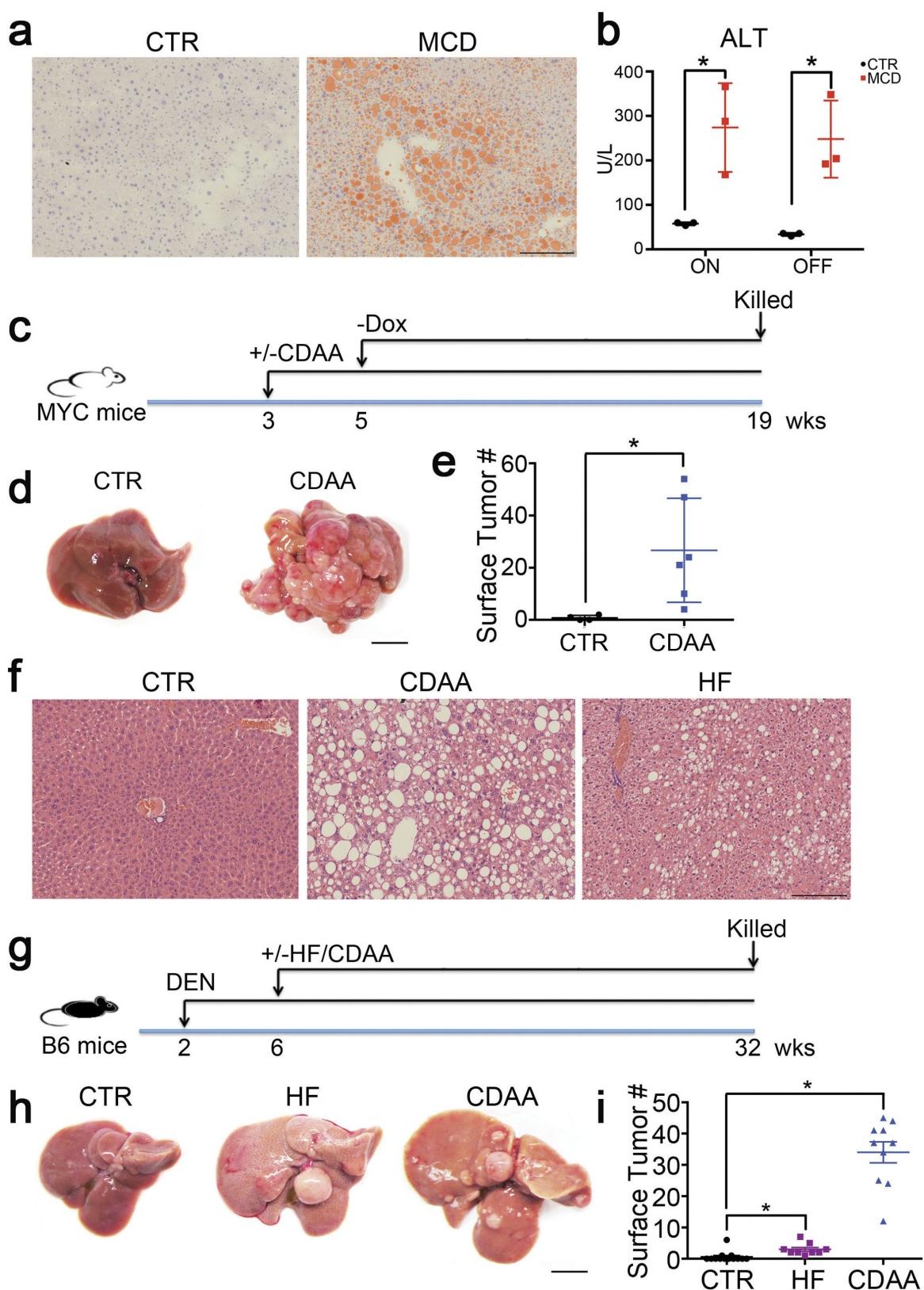
Oxygen consumption assay. OCR was measured using an XFe96 Extracellular Flux Analyzer (Seahorse Bioscience) as previously described³⁸. AutoMACS-sorted mouse CD4⁺ and CD8⁺ T lymphocytes were attached to XFe96 cell culture plates using Cell-Tak (BD Bioscience) in RPMI media with 11 mM glucose. Cells were activated with 1:1 CD3:CD28 beads (Miltenyi BioTech) and vehicle or 50 μM C18:2 was added. Twenty-four hours after activation, cells were incubated in serum-free XF Base Media (Seahorse Bioscience) supplemented with 10 mM glucose, 2 mM pyruvate and 2 μM glutamine, pH 7.4, along with 50 μM C18:2 if previously present, for 30 min at 37 °C in a CO₂-free cell culture incubator before beginning the assay. Five consecutive measurements, each representing the mean of 8 wells, were obtained at baseline and after sequential addition of 1.25 μM oligomycin, 0.25 μM trifluorocarbonylcyanide phenylhydrazone (FCCP), and 1 μM each of rotenone and antimycin A (all drugs from Seahorse Bioscience). OCR values were normalized to cell number as measured by the CyQUANT Cell Proliferation Assay Kit (Life Technologies).

Human studies. Human liver samples were stained as previously described⁸. For immunostaining, formalin-fixed, paraffin-embedded human liver tissue samples were retrieved from the archives of the Institute of Surgical Pathology, University Hospital Zurich. Fibrosis grade was analysed for NASH according to NAFLD activity score (NAS)³⁹ and for others according to METAVIR score⁴⁰. The study was approved by the local ethics committee (Kantonale Ethikkommission Zürich, application number KEK-ZH-Nr. 2013-0382). Human PBMCs from healthy donors

were obtained on an NIH-approved protocol and prepared as described previously⁴¹. Informed consent was obtained from all subjects.

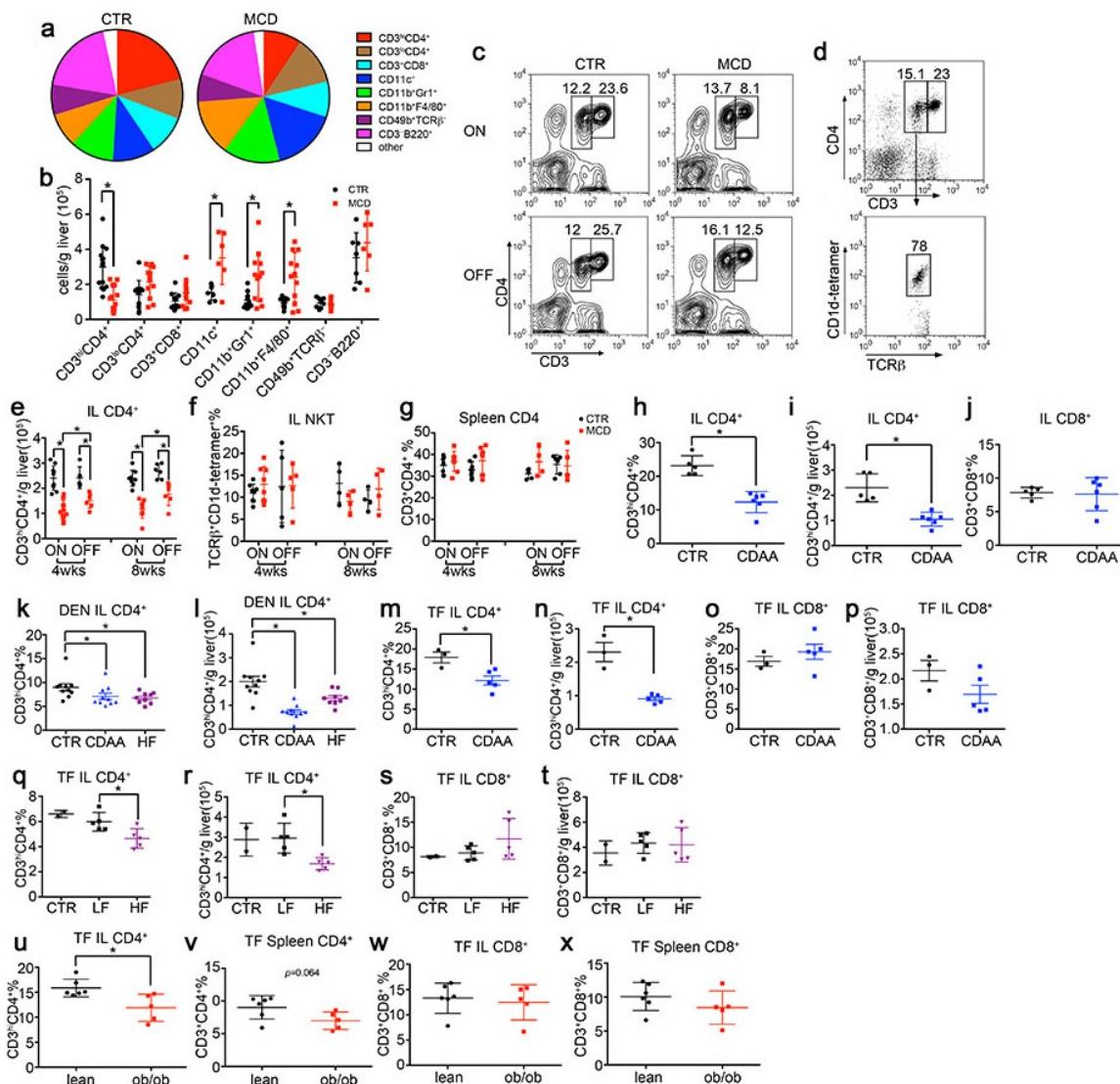
Statistical analysis. The sample sizes for animal studies were guided by a previous study in our laboratory in which the same MYC transgenic mouse strain was used. No animals were excluded. Neither randomization nor blinding were done during the *in vivo* study. However, mice from the same littermates were evenly distributed into control or treatment groups whenever possible. The sample size for the patient studies was guided by a recent publication also studying NASH-induced HCC, but focused on different aspects⁸. Statistical analysis was performed with GraphPad Prism 6 (GraphPad Software). Significance of the difference between groups was calculated by Student's unpaired *t*-test, one-way or two-way ANOVA (Tukey's and Bonferroni's multiple comparison test). Welch's corrections were used when variances between groups were unequal. *P* < 0.05 was considered as statistically significant.

31. Andersson, J. *et al.* CD4⁺ FoxP3⁺ regulatory T cells confer infectious tolerance in a TGF-β-dependent manner. *J. Exp. Med.* **205**, 1975–1981 (2008).
32. Nakagawa, H. *et al.* ER stress cooperates with hypernutrition to trigger TNF-dependent spontaneous HCC development. *Cancer Cell* **26**, 331–343 (2014).
33. Ikehara, Y. *et al.* CD4⁺ Vo.14 natural killer T cells are essential for acceptance of rat islet xenografts in mice. *J. Clin. Invest.* **105**, 1761–1767 (2000).
34. Hanczko, R. *et al.* Prevention of hepatocarcinogenesis and increased susceptibility to acetaminophen-induced liver failure in transaldolase-deficient mice by N-acetylcysteine. *J. Clin. Invest.* **119**, 1546–1557 (2009).
35. Radaeva, S. *et al.* Interferon-α activates multiple STAT signals and down-regulates c-Met in primary human hepatocytes. *Gastroenterology* **122**, 1020–1034 (2002).
36. Chen, R. F. Removal of fatty acids from serum albumin by charcoal treatment. *J. Biol. Chem.* **242**, 173–181 (1967).
37. Huynh, F. K., Green, M. F., Kovacs, T. R. & Hirschey, M. D. Measurement of fatty acid oxidation rates in animal tissues and cell lines. *Methods Enzymol.* **542**, 391–405 (2014).
38. Pelletier, M., Billingham, L. K., Ramaswamy, M. & Siegel, R. M. Extracellular flux analysis to monitor glycolytic rates and mitochondrial oxygen consumption. *Methods Enzymol.* **542**, 125–149 (2014).
39. Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
40. Bedossa, P. & Poupon, T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* **24**, 289–293 (1996).
41. Hoechst, B. *et al.* A new population of myeloid-derived suppressor cells in hepatocellular carcinoma patients induces CD4⁺CD25⁺Foxp3⁺ T cells. *Gastroenterology* **135**, 234–243 (2008).



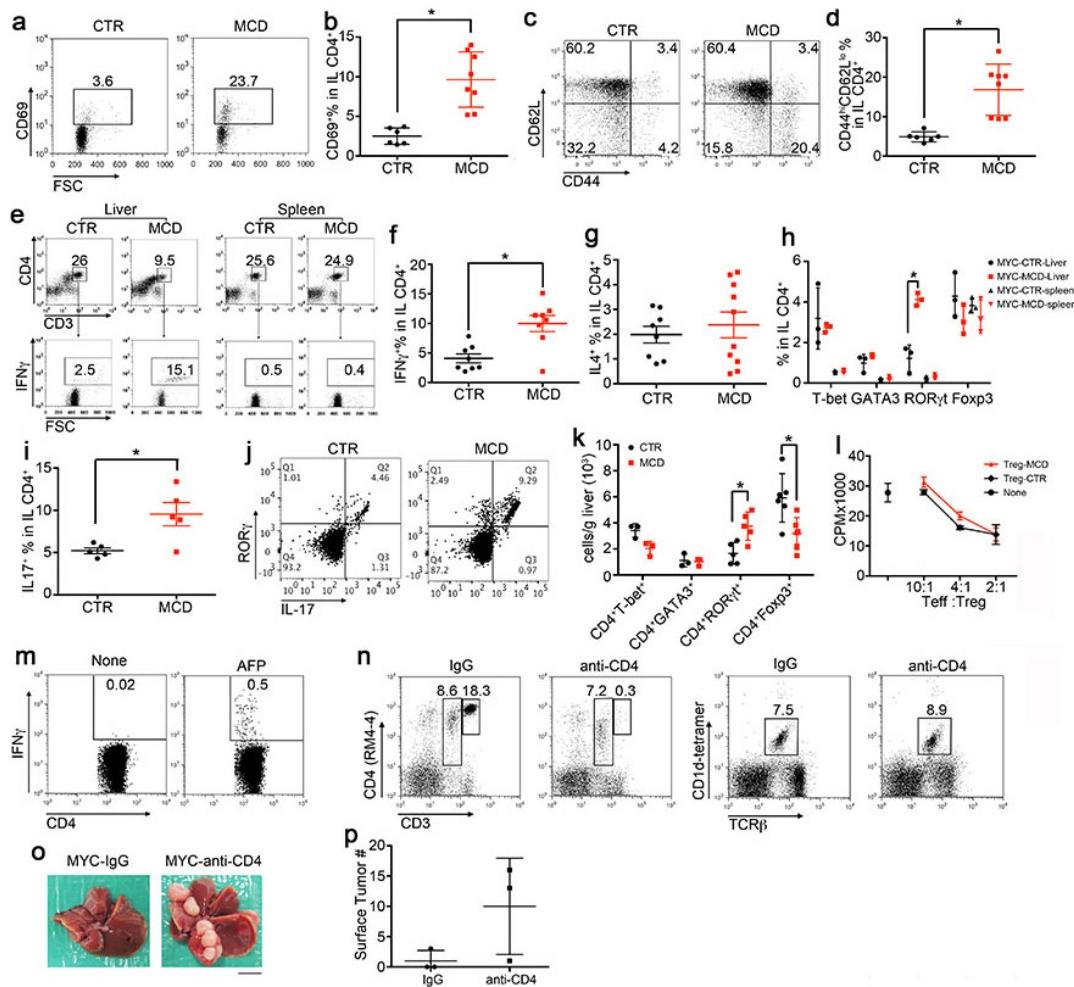
Extended Data Figure 1 | MCD, CDAA and HF diets induce NAFLD and promote HCC. **a**, Representative imagines of Oil Red O staining of MYC-ON mice fed MCD or CTR. Scale bar, 100 μ m. **b**, Serum ALT levels analysis. Data are mean \pm s.e.m.; $n = 4$, * $P < 0.05$, one-way ANOVA. **c–e**, The effect of the CDAA diet on tumour development in MYC transgenic mice. Experimental setup, representative liver images and liver surface tumour counts are shown. Scale bar, 10 mm. Data are

mean \pm s.e.m.; $n = 6$ for CDAA and $n = 5$ mice for CTR, $P = 0.0345$, Student's t -test. **f-i**, The effect of the CDAA and HF diet on liver carcinogenesis in diethylnitrosamine (DEN)-injected C57BL/6 mice. Experimental setup, representative tumour-free H&E stainings, macroscopic liver images and surface tumour counts are shown. Scale bar, 100 μm . Data are mean \pm s.e.m.; $n = 13$ for CTR, $n = 9$ for HF, $n = 10$ for CDAA, $*P < 0.05$, one-way ANOVA.



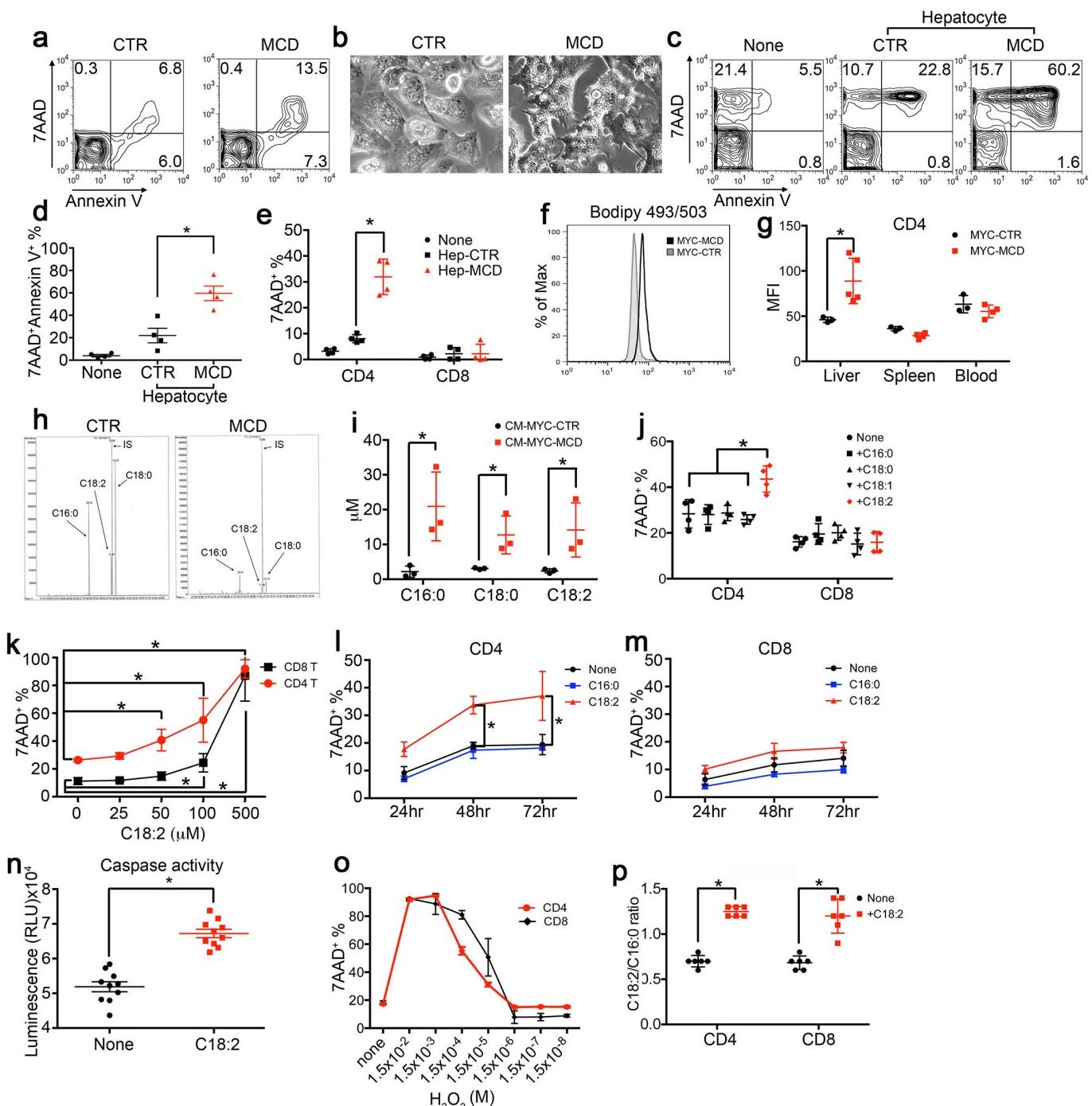
Extended Data Figure 2 | Immune cell monitoring in NAFLD-HCC.
a–j, MYC mice were fed with an MCD diet or CTR diet. **a, b**, Intrahepatic immune cells were determined by flow cytometry. Composition (a) and absolute numbers (b) of different intrahepatic immune cell subsets in MYC-ON mice, which were kept for 4 weeks on an MCD diet or CTR diet. Data are mean \pm s.e.m.; $n \geq 6$, $*P < 0.05$, one-way ANOVA. **c**, Representative contour plots of intrahepatic CD4 $^{+}$ T lymphocytes. **d**, Representative dot plots of CD1d-tetramer staining in CD3 lo CD4 $^{+}$ population. **e–g**, Absolute number of intrahepatic CD4 $^{+}$ T lymphocytes, frequencies of NK T cells and splenic CD4 $^{+}$ T lymphocytes were measured by flow cytometry. Data are mean \pm s.e.m.; $n = 4$, $*P < 0.05$, two-way ANOVA. **h–j**, Intrahepatic CD4 $^{+}$ and CD8 $^{+}$ T lymphocyte levels in MYC-ON mice fed with a CDAA diet for 16 weeks. Data are mean \pm s.e.m.; $n = 6$ for CDAA and $n = 5$ for CTR, $*P < 0.05$, Student's

t-test. **k, l**, Intrahepatic CD4 $^{+}$ T lymphocyte levels in DEN-injected BL/6 male mice treated with a CDAA diet, HF diet or CTR for 7 months. Data are mean \pm s.e.m.; $n = 13$ for CTR, $n = 9$ for HF, $n = 10$ for CDAA, $*P < 0.05$, one-way ANOVA. **m–p**, Intrahepatic CD4 $^{+}$ and CD8 $^{+}$ T lymphocytes in tumour-free C57BL/6 mice treated with a CDAA diet for 16 weeks. TF, tumour free. Data are mean \pm s.e.m.; $n = 3$ for CTR, $n = 5$ for CDAA, $*P < 0.05$, Student's *t*-test. **q–t**, Intrahepatic CD4 $^{+}$ and CD8 $^{+}$ T lymphocytes in tumour-free C57BL/6 mice treated with an HF or low-fat (LF) diet for 6 months. Data are mean \pm s.e.m.; $n = 2$ for CTR, $n = 5$ for LF, $n = 5$ for HF, $*P < 0.05$, one-way ANOVA. **u–x**, CD4 $^{+}$ and CD8 $^{+}$ T lymphocytes in 12-week-old male *ob/ob* or wild-type lean mice. Data are mean \pm s.e.m., $n = 5$, $*P < 0.05$, Student's *t*-test. **w, x**, MYC mice were fed with MCD or CTR. Macrophage and CD11b $^{+}$ Gr1 $^{+}$ populations were measured. Data are mean \pm s.e.m.; $n \geq 4$, $*P < 0.05$, two-way ANOVA.



Extended Data Figure 3 | Intrahepatic CD4⁺ lymphocytes are activated in NAFLD, and CD4 depletion enhances HCC. **a–k**, MYC-ON mice were fed with MCD or CTR for 4 weeks. **a–d**, CD69 and CD44^{hi}CD62L^{lo} subsets in intrahepatic CD4⁺ T lymphocytes were measured. Data are mean ± s.e.m.; n = 8 for MCD and n = 6 for CTR, *P < 0.05, Student's t-test. **e–g**, *Ex vivo* IFN- γ , IL-4 production in intrahepatic CD4⁺ T lymphocytes were determined. Data are mean ± s.e.m.; n = 8, *P < 0.05, Student's t-test. **h**, *Ex vivo* staining of T-bet, GATA3, ROR- γ t and Foxp3 levels in intrahepatic and splenic CD4⁺ T lymphocytes. Data are mean ± s.e.m.; n = 3, *P < 0.05, two-way ANOVA. **i**, *Ex vivo* IL-17 production by intrahepatic CD4⁺ T lymphocytes. Data are mean ± s.e.m.; n = 5, *P < 0.05, Student's t-test. **j**, Representative dot plots of ROR- γ t/

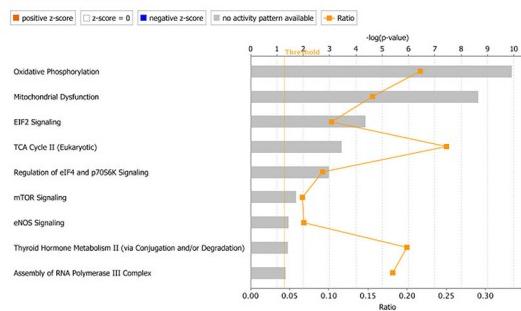
IL-17 staining in intrahepatic CD4⁺ T lymphocytes. **k**, Absolute number of intrahepatic CD4⁺ lymphocyte subsets. Data are mean ± s.e.m.; n = 3, *P < 0.05, two-way ANOVA. **l**, Suppressive function assay of isolated hepatic T_{reg} cells from Foxp3-GFP mice kept on MCD or CTR for 4 weeks. **m**, Detection of AFP-specific CD4⁺ T lymphocytes in spleen from MYC-MCD mice. **n**, Selective depletion of intrahepatic CD4⁺ T lymphocytes but not NK T cells by i.p. injection of 50 µg anti-CD4 antibody (clone GK1.5). **o**, **p**, MYC-ON mice on CTR received 50 µg of GK1.5 antibody or isotype control i.p. once per week for 8 weeks. Representative liver imagines and surface tumour counts are shown. Scale bar, 10 mm. Data are mean ± s.e.m., n = 3.



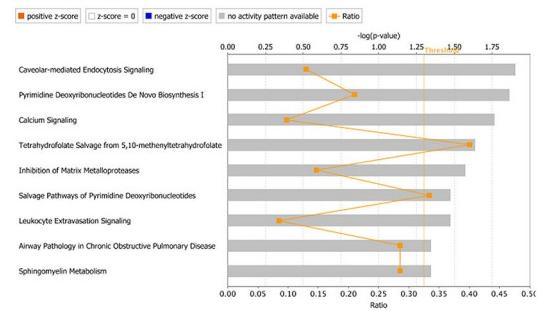
Extended Data Figure 4 | Lipid-laden hepatocytes release C18:2 and induce CD4⁺ T lymphocyte death via apoptosis. **a**, Representative contour plots of *ex vivo* 7AAD/annexin V staining of intrahepatic CD4⁺ T lymphocytes from MYC-ON mice fed with MCD or CTR. **b**, Representative phase-contrast images of primary hepatocytes from MYC-ON mice after MCD or CTR treatment. **c–e**, Isolated primary hepatocytes from MYC-ON mice on MCD or CTR were cocultured with isolated CD4⁺ T lymphocytes or splenocytes. Cell death levels were measured by flow cytometry. Data are mean \pm s.e.m.; $n = 4$, one-way or two-way ANOVA. **f, g**, BODIPY 493/503 staining of CD4⁺ T lymphocytes in liver, spleen or blood from MYC-ON mice with MCD or CTR. Data are mean \pm s.e.m.; $n = 4$, * $P < 0.05$, two-way ANOVA. **h, i**, Identification

of FFAs in hepatocyte conditioned medium by gas chromatography/mass spectrometry (GC/MS). Data are mean \pm s.e.m.; $n = 3$, * $P < 0.05$, two-way ANOVA. **j**, Anti-CD3/28 bead-activated splenocytes were treated with different FFAs, and cell death level in CD4⁺ or CD8⁺ T lymphocytes was determined. Data are mean \pm s.e.m.; $n = 4$, * $P < 0.05$, two-way ANOVA. **k–m**, Dose-response curve and time course of C18:2-induced cell death in CD4⁺ or CD8⁺ T lymphocytes. **n**, Caspase3/7 activity in CD4⁺ T lymphocytes after C18:2 treatment. Data are mean \pm s.e.m.; $n = 9$, * $P < 0.05$, Student's *t*-test. **o**, Dose-response curve of H₂O₂-induced cell death in CD4⁺ or CD8⁺ T lymphocytes. **p**, Uptake of C18:2 by CD4⁺ and CD8⁺ T lymphocytes after incubation with 50 μM C18:2 for 2 h. Data are mean \pm s.e.m.; $n = 6$, * $P < 0.05$, two-way ANOVA.

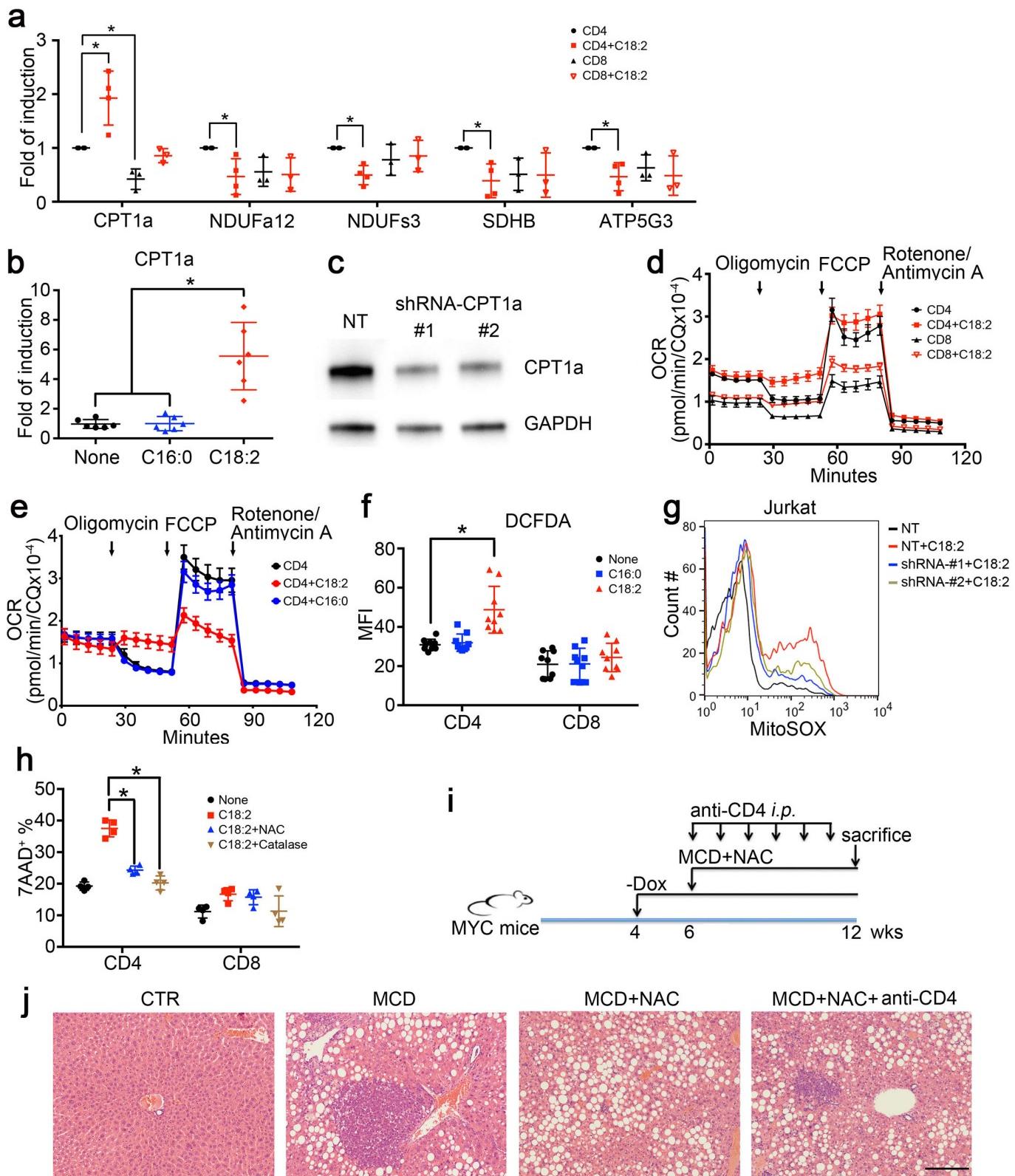
CD4+LA vs CD4



CD8+LA vs CD8

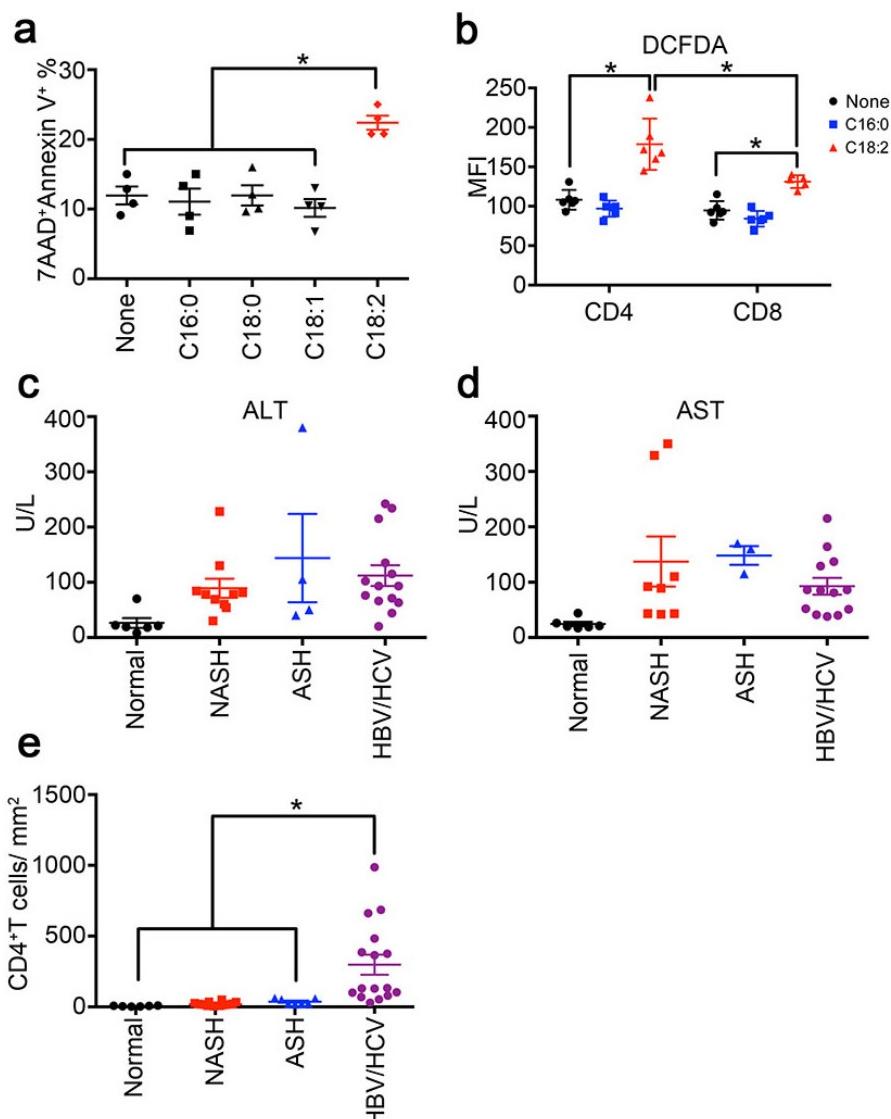


Extended Data Figure 5 | Ingenuity pathway analysis of microarray data. CD4⁺ and CD8⁺ T lymphocytes sorted from C18:2-treated splenocytes were subjected to microarray analysis. Pathway analysis was done by ingenuity pathway analysis (IPA). $n = 3$. Ratio is the number of changed genes divided by total genes in the pathway.



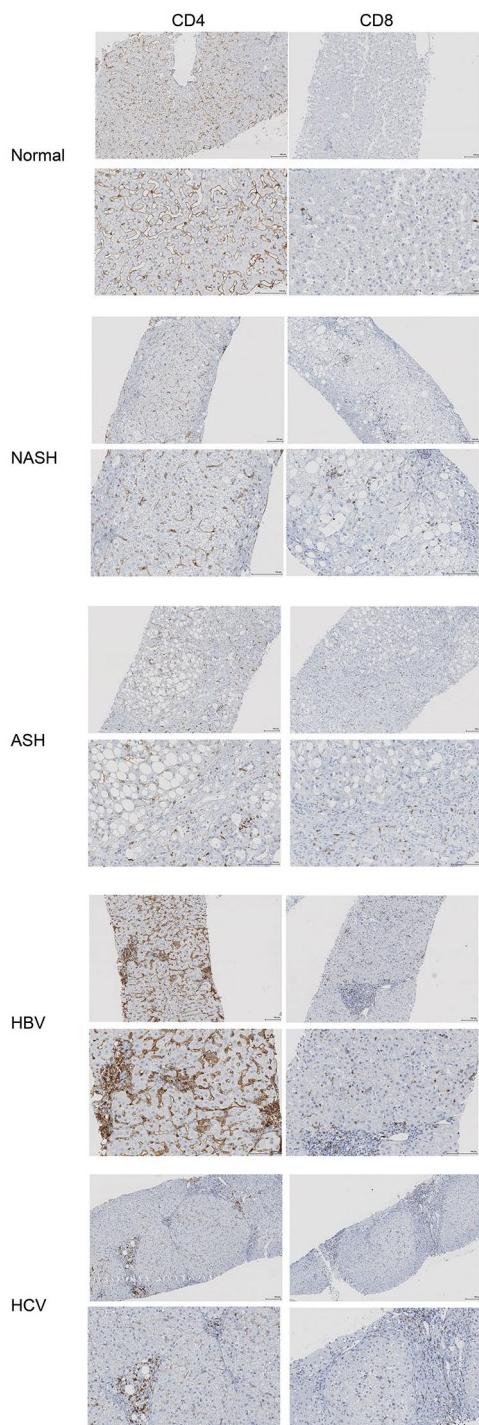
Extended Data Figure 6 | Mitochondrial ROS mediates C18:2-induced CD4⁺ T lymphocyte death *in vitro* and *in vivo*. **a**, Real-time polymerase chain reaction (PCR) confirmed the gene changes from microarray. Data are mean \pm s.e.m.; $n = 3$, $*P < 0.05$, two-way ANOVA. **b**, *Cpt1a* mRNA level in Jurkat cells after FFA treatment. Data are mean \pm SEM; $n = 6$, $*P < 0.05$, one-way ANOVA. **c**, Expression of CPT1a in wild-type and two knockdown Jurkat cells. NT, non-targeting control. **d, e**, OCR analysis of activated CD4⁺ and CD8⁺ T lymphocytes upon C18:2 or C16:0 incubation. **f**, ROS levels of CD4⁺ or CD8⁺ T lymphocytes in splenocytes

treated with C18:2 or C16:0. Data are mean \pm s.e.m.; $n = 8$, $*P < 0.05$, two-way ANOVA. **g**, Mitochondrial ROS in wild-type and two CPT1 knockdown Jurkat cells. **i**, Cell death of CD4⁺ or CD8⁺ T lymphocytes in splenocytes treated with C18:2 in the presence of NAC or catalase. Data are mean \pm s.e.m.; $n = 4$, $*P < 0.05$, two-way ANOVA. **i, j**, *In vivo* blocking ROS with NAC in MYC-ON mice treated with MCD. Some mice also received CD4 antibody depletion. Experimental setup and representative H&E liver sections are shown. Scale bar, 200 μ m.



Extended Data Figure 7 | C18:2 induces cell death in human CD4⁺ T lymphocytes, and NASH patients have lower intrahepatic CD4⁺ T lymphocytes. **a**, Cell death levels of sorted human CD4⁺ T lymphocytes treated with different FFAs. Data are mean \pm s.e.m.; $n=4$, * $P<0.05$, one-way ANOVA. **b**, ROS level of CD4⁺ or CD8⁺ T lymphocyte in peripheral blood mononuclear cells (PBMCs) treated with C18:2 or C16:0.

Data are mean \pm s.e.m.; $n=6$, * $P<0.05$, two-way ANOVA. **c, d**, Serum ALT and AST concentration in different patients. **e**, Intrahepatic CD4⁺ T lymphocyte count in biopsies. CD4⁺ T lymphocytes were identified by immunohistochemistry. Data are mean \pm s.e.m.; normal = 6, NASH = 16, ASH = 8, HBV/HCV = 16, * $P<0.05$, one-way ANOVA.



Extended Data Figure 8 | Immunohistochemistry staining of intrahepatic CD4⁺ or CD8⁺ T lymphocytes in patient biopsies.

Representative CD4 or CD8 immunohistochemistry images of liver biopsies from healthy individuals, NASH, ASH patients or patients with HBV or HCV. For each condition, two different magnifications are shown. Scale bar, 100 μm.

Extended Data Table 1 | Fatty acid composition of hepatic lipids

Fatty Acid	ON-CTR	ON-MCD	OFF-CTR	OFF-MCD
C12:0	17.2 ± 4.2	16.7 ± 2.1	10.3 ± 2.2	18.6 ± 2.6
C14:0	143.9 ± 8.1	204.8 ± 21.1	195.2 ± 34.4	237.6 ± 25.7
C15:0	86.9 ± 6.1	62.2 ± 6.1	77.9 ± 4.2	53.1 ± 4.0
C16:0	10986.0 ± 703.0	7576.4 ± 486.5 ^a	11325.9 ± 517.8	8274.1 ± 543.9 ^b
C16:1	467.2 ± 49.0	88.5 ± 7.7	430.9 ± 52.6	169.2 ± 17.7
C17:0	216.1 ± 18.4	199.0 ± 17.2	249.8 ± 17.0	171.7 ± 12.9
C17:1	0.5 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
C18:0	6861.1 ± 520.0	7228.8 ± 463.1	8335.6 ± 287.9	7700.4 ± 557.0
C18:1	1498.2 ± 126.0	1432.8 ± 108.1	1545.8 ± 102.3	1773.6 ± 111.2
C18:2	4664.8 ± 464.5	6492.8 ± 356.1 ^a	4516.3 ± 582.9	7584.1 ± 462.3 ^b
C18:3N3	30.0 ± 3.8	183.5 ± 23.3	28.3 ± 4.0	187.7 ± 20.1
C18:3N6	71.0 ± 8.0	37.8 ± 2.4	65.1 ± 8.0	48.5 ± 2.8
C18:4	2.3 ± 0.4	3.7 ± 0.5	1.4 ± 0.3	4.1 ± 0.4
C20:0	80.8 ± 6.8	29.0 ± 2.4	16.4 ± 1.1	13.2 ± 1.3
C20:1	2.5 ± 0.3	3.1 ± 0.4	1.1 ± 0.2	1.8 ± 0.2
C20:2	119.7 ± 8.7	183.0 ± 14.3	108.3 ± 4.3	158.3 ± 16.7
C20:3N3	53.0 ± 4.0	38.0 ± 5.6	53.8 ± 5.7	24.1 ± 3.8
C20:3N6	973.8 ± 28.6	592.2 ± 75.3	709.6 ± 33.2	529.5 ± 58.0
C20:3N9	13.5 ± 1.2	3.4 ± 0.3	10.0 ± 0.8	2.2 ± 0.3
C20:4	10503.7 ± 751.6	10949.5 ± 799.5	14287.7 ± 619.5	10801.1 ± 1218.5 ^b
C20:5	118.6 ± 9.7	25.0 ± 2.1	103.3 ± 10.3	28.5 ± 2.0
C22:0	19.8 ± 1.6	6.5 ± 0.7	7.8 ± 1.0	3.2 ± 0.2
C22:1	20.3 ± 1.6	5.5 ± 0.5	6.1 ± 0.3	2.2 ± 0.3
C22:2	7.8 ± 0.9	2.5 ± 0.3	0.2 ± 0.1	0.0 ± 0.0
C22:3	2.4 ± 0.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
C22:4	140.1 ± 8.3	419.9 ± 42.2	136.7 ± 4.3	296.4 ± 37.5
C22:5N3	454.8 ± 29.2	997.8 ± 76.6	523.1 ± 23.4	706.5 ± 114.3
C22:5N6	336.1 ± 29.8	80.0 ± 6.8	206.1 ± 7.6	87.7 ± 11.4
C22:6	9237.6 ± 687.1	5207.0 ± 388.0 ^a	10704.6 ± 426.9	6414.9 ± 638.9 ^b
C23:0	1.2 ± 0.2	0.2 ± 0.1	0.2 ± 0.1	0.0 ± 0.0
C24:0	6.5 ± 0.4	2.9 ± 0.5	3.3 ± 0.4	1.8 ± 0.1
C24:1	1.1 ± 0.1	0.4 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Total	47313.0 ± 3047.1	41742.3 ± 2471.9	53664.5 ± 1953.3	45295.3 ± 3554.2

Mouse livers from MYC-ON or MYC-OFF mice receiving 4 weeks of MCD or CTR diet treatment were subjected to total fatty acid profiling by GC/MS as described in Methods. Results (in pmol per mg tissue) are expressed as mean ± s.e.m. ($n=6$ per group).

* $P<0.05$ for ON-CTR versus ON-MCD, two-way ANOVA.

† $P<0.05$ for OFF-CTR versus OFF-MCD, two-way ANOVA.

Extended Data Table 2 | Overview of patient cohort for immunohistochemistry analysis

No	Disease	Age	Gender	ALT(U/L)	AST(U/L)	Fibrosis	Cirrhosis	Activity
1	Normal	39	M	70	44	0	N/A	N/A
2	Normal	29	F	18	26	0	N/A	N/A
3	Normal	50	F	8	19	0	N/A	N/A
4	Normal	44	M	21	20	0	N/A	N/A
5	Normal	26	M	19	17	0	N/A	N/A
6	Normal	56	F	22	21	0	N/A	N/A
7	NASH	55	M	78	43	4	1a	N/A
8	NASH	45	F	30	92	5	3	N/A
9	NASH	61	F	54	89	4	2	N/A
10	NASH	41	M	81	x	4	x	N/A
11	NASH	43	M	x	x	4	1a	N/A
12	NASH	35	M	x	x	4	1b	N/A
13	NASH	27	M	78	x	4	2	N/A
14	NASH	69	M	x	x	4	3	N/A
15	NASH	39	F	69	350	5	1b	N/A
16	NASH	54	F	130	110	4	2	N/A
17	NASH	59	M	228	329	6	4	N/A
18	NASH	49	M	x	x	5	2	N/A
19	NASH	27	F	x	x	5	2	N/A
20	NASH	44	M	84	42	5	1b	N/A
21	NASH	64	F	60	43	4	2	N/A
22	NASH	50	F	x	x	4	1a	N/A
23	ASH	63	F	40	115	4	N/A	N/A
24	ASH	51	M	x	x	2	N/A	N/A
25	ASH	75	F	x	x	3	N/A	N/A
26	ASH	46	M	50	160	4	N/A	N/A
27	ASH	58	M	105	x	4	N/A	N/A
28	ASH	70	M	x	x	1	N/A	N/A
29	ASH	43	M	380	170	0	N/A	N/A
30	ASH	64	M	x	x	3	N/A	N/A
31	HBV	19	M	234	86	1	N/A	2
32	HBV	60	M	115	86	3	N/A	1
33	HBV	41	F	76	52	1	N/A	2
34	HBV	27	F	71	41	1	N/A	2
35	HBV	48	F	x	x	1	N/A	1
36	HBV	77	M	63	137	1	N/A	1
37	HCV	54	F	44	38	3	N/A	2
38	HCV	50	M	102	85	1	N/A	1
39	HCV	58	M	66	x	2	N/A	1
40	HCV	57	F	215	164	3	N/A	1
41	HCV	55	M	135	129	2	N/A	2
42	HCV	73	F	93	81	3	N/A	1
43	HCV	71	F	x	x	4	N/A	2
44	HCV	60	F	20	40	4	N/A	2
45	HCV	70	F	242	215	4	N/A	2
46	HCV	57	F	93	51	2	N/A	2

Fibrosis grade was analysed for NASH according to NAFLD activity score (NAS) and for others according to METAVIR score. Disease activity of viral hepatitis was analysed according to METAVIR score. N/A, not applicable; x, not available.

Structure, inhibition and regulation of two-pore channel TPC1 from *Arabidopsis thaliana*

Alexander F. Kintzer¹ & Robert M. Stroud¹

Two-pore channels (TPCs) comprise a subfamily (TPC1–3) of eukaryotic voltage- and ligand-gated cation channels^{1,2} with two non-equivalent tandem pore-forming subunits that dimerize to form quasi-tetramers. Found in vacuolar³ or endolysosomal⁴ membranes, they regulate the conductance of sodium⁵ and calcium^{3,6} ions, intravesicular pH⁵, trafficking⁷ and excitability^{8,9}. TPCs are activated by a decrease in transmembrane potential^{1,3,9,10} and an increase in cytosolic calcium concentrations^{1,10}, are inhibited by low luminal pH and calcium¹¹, and are regulated by phosphorylation^{12,13}. Here we report the crystal structure of TPC1 from *Arabidopsis thaliana* at 2.87 Å resolution as a basis for understanding ion permeation^{3,4,10}, channel activation^{1,5,10}, the location of voltage-sensing domains^{1,9,10} and regulatory ion-binding sites^{11,14}. We determined sites of phosphorylation^{3,4} in the amino-terminal and carboxy-terminal domains that are positioned to allosterically modulate cytoplasmic Ca²⁺ activation. One of the two voltage-sensing domains (VSD2) encodes voltage sensitivity and inhibition by luminal Ca²⁺ and adopts a conformation distinct from the activated state observed in structures of other voltage-gated ion channels^{15,16}. The structure shows that potent pharmacophore *trans*-Ned-19 (ref. 17) acts allosterically by clamping the pore domains to VSD2. In animals, Ned-19 prevents infection by Ebola virus and other filoviruses, presumably by altering their fusion with the endolysosome and delivery of their contents into the cytoplasm⁷.

Diffract was optimized and the final conditions depended on relipidation, partial dehydration, Ned-19 action (ref. 17), and deletion of residues 2–11. The structure was determined *de novo* to 2.87 Å resolution by nine metal substitutions and derivatives, and refined to $R_{\text{cryst}} = 29.7\%$ and $R_{\text{free}} = 33.9\%$ (Methods and Extended Data Tables 1–3).

TPC1 consists of two non-identical *Shaker*-like^{18,19} pore-forming subunits, domain 1 and 2, separated by an EF-hand domain (Fig. 1). Each subunit contains 12 transmembrane helices (S1–S12). Two central pore domains, P1 (S5–S6) and P2 (S11–S12), couple the voltage-sensing domains VSD1 (S1–S4) and VSD2 (S7–S10) and modulatory cytosolic N-terminal domain (NTD), the central EF-hand (EF) domain, and C-terminal domain (CTD), which extend 20 Å into the cytoplasm (Fig. 2, Extended Data Fig. 1, see Methods for detailed domain assignments). The experimentally determined electron density unbiased by any model (Extended Data Fig. 2), shows that TPC1 forms a rectangular structure, ~100 Å × 70 Å in the membrane plane (Fig. 1). While our manuscript was under review, we became aware of a study reporting the crystal structure of the same molecule, accompanied by extensive functional characterizations¹⁰, which is generally in good agreement with our results. We additionally observe the CTD, a lipid and inhibitor Ned-19, and there is some variation in observed ion-binding sites. We discuss our results in light of the observations in ref. 10.

Plant and human TPCs respond to phosphoregulation by specific kinases^{3,4}. We determined sites of phosphorylation in TPC1 using a combination of mass spectrometry and binding of Pro-Q Diamond phosphoprotein stain to N- and C-terminal truncations (Extended Data Fig. 3 and Supplementary Discussion). Residues S22, T26, and T29 in

the NTD are sites of phosphorylation in TPC1. S22 is conserved in both human (h)TPCs, whereas T26 is only found in hTPC1 (Extended Data Fig. 1). The ordered structure begins with NTD helix 2 (NTDh2; residues P30–G46). NTDh2 makes hydrophobic and polar contacts with

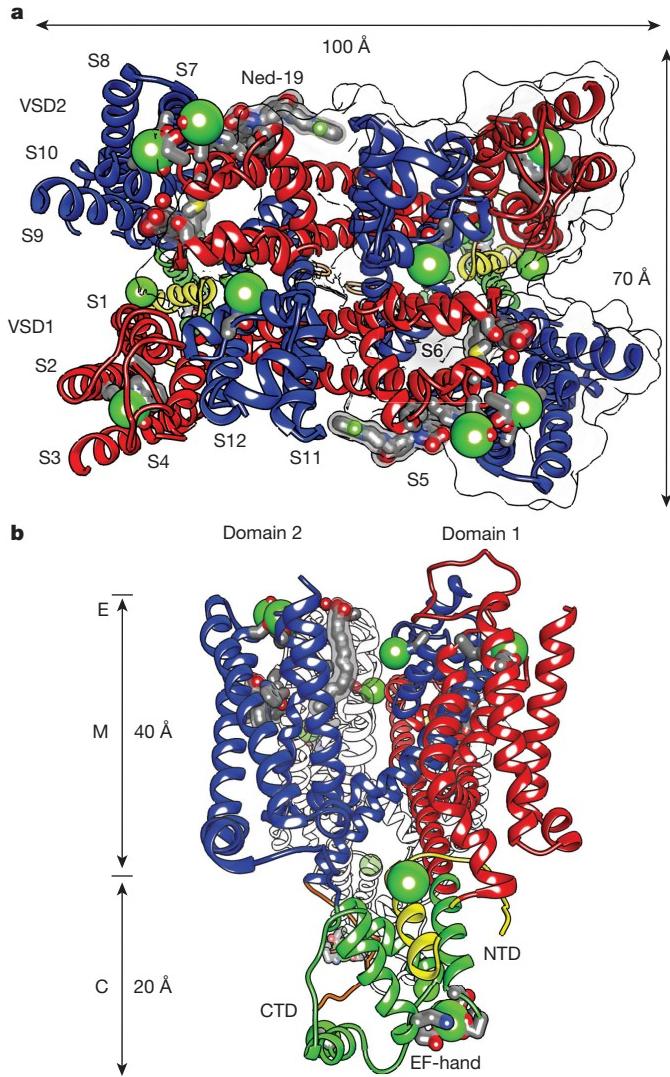


Figure 1 | Overview of the TPC1 Structure. **a**, Top view of the TPC1 structure from the luminal side onto the membrane plane. **b**, Side view of the TPC1 structure from the right side with the vertical direction representing the perpendicular to the membrane plane. Domains and transmembrane helices are labelled. E, M, and C denote approximate endolysosomal, membrane and cytosolic boundaries, respectively. The positions of Ca²⁺ ion-binding sites (green spheres), bound lipids and Ned-19 are shown. Approximate geometric dimensions of TPC1 are indicated.

¹Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA.

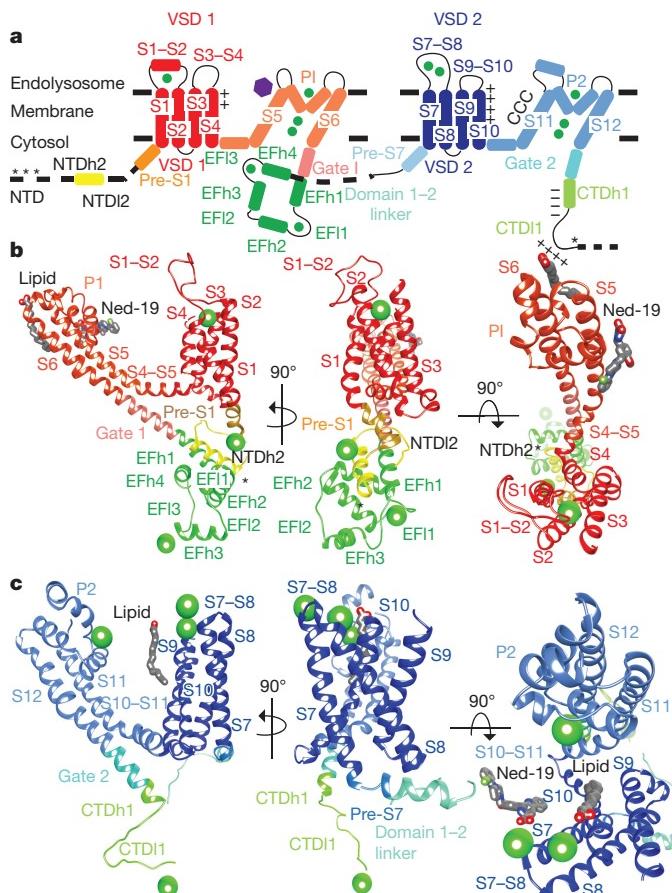


Figure 2 | Structural details of the TPC1 monomer. Structure and domain boundaries in TPC1. **a**, The diagram shows the arrangement of structural domains, coloured to match the structures below. Helices are depicted as cylinders and loops as lines. **b, c**, Structural details of TPC1 domain 1 (**b**) and domain 2 (**c**) from three perpendicular views. Asterisks indicate phosphorylation sites. The charges in S4 (++)+, S10 (++++)+, poly-E (—), in CTDh1, poly-R (++++) in CTDl1, and CXXCXC (CCC) are shown. Ca²⁺ ions are shown in green, and the Ned-19 binding site is shown as a purple hexagon.

the EF-hand domain helix 1 (EFh1), EFh2, EFh4, and VSD1 (S2–S3), leading into a NTD loop 2 (NTDl2) that makes an almost 180° turn to connect to the pre-S1 helix of VSD1. The proximity of the NTD phosphorylation sites to the Ca²⁺-activation site in the EF-hand^{10,14}, EFh3 and loop EFl3, and VSD1 suggests that phosphorylation in the NTD could modulate channel opening by influencing the local structure of the Ca²⁺ site or conformation of VSD1 via a salt bridge (E50–R200) between NTDl2 and VSD1 (S4–S5).

Plant and human TPC1s form voltage-dependent inwardly rectifying (towards the cytosol) ion channels, whereas hTPC2 is voltage insensitive^{1,5,9,10,12}. In contrast to voltage-gated Ca_v, Na_v and K_v channels, activation of TPC channels is characteristically much slower and they are not subject to inactivation, hence the original naming of plant TPCs as slow vacuolar channels^{1,5,9,10,12}.

Voltage sensing is primarily mediated by VSD2, but not VSD1 (ref. 10; Fig. 3). VSD2 contains a classic voltage-sensing motif with four arginines, R531, R537, R540 and R543, in S10 (Fig. 3b, Extended Data Fig. 4a) and a conserved counter anion charge-transfer centre (Y475–E478–R537) in S8 (Extended Data Fig. 4b), as in Na_vAb¹⁶. Three other arginines form corresponding ion pairs with anions on S9 (E511–R531, D500–R540, and E494–R543). In ref. 10, it is shown that the three arginines R537, R540 and R543 are required for voltage sensing in TPC1. The second voltage-sensing arginine R537 interacts with the conserved luminal charge-transfer centre near Y475 (refs 10, 16), creating an opening in the cytosolic side of VSD2, and exposing R543

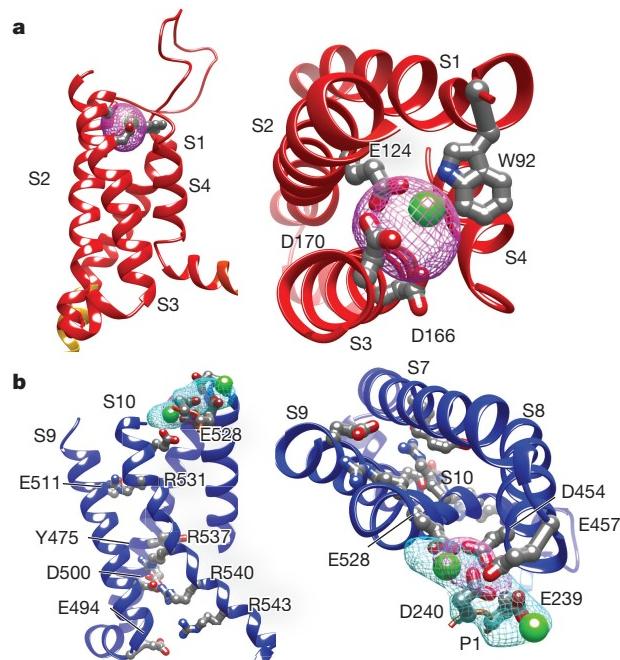


Figure 3 | Voltage dependence and its modulation by ions. **a, b**, Side (left) and top (right) views of Ca²⁺-binding sites in VSD1 (**a**) and VSD2 (**b**). Ca²⁺ ions are shown in green. Ion-coordinating residues are labelled in VSD1 and 2. Voltage-sensing residues are shown in VSD2. Ba²⁺ (cyan) and Yb³⁺ (magenta) isomorphous difference density peaks and atom positions contoured at 10 σ , numbered according to Extended Data Fig. 2c.

to the cytosol. VSD2 is in a novel conformation with respect to other voltage-gated ion channel structures, which each have the last charge, equivalent to R543, situated at the charge-transfer centre, thought to represent activated VSDs^{15,16}. The luminal/extracellular side of activated VSDs adopts a more open conformation, whereas the cytosolic side is narrow and closed. This may mean that the TPC1 VSD2 structure represents an inactive state. VSD2 is also indicated as the functional voltage sensor in human TPC1, as the R539I mutation (corresponding to R537 in plant TPC1 VSD2) eliminates most of the observed voltage sensitivity⁹. In the human TPC2 channel the corresponding residue is I551, consistent with its lack of voltage sensitivity.

VSD1 is structurally divergent from VSD2, containing only two arginines (R185 and R191) in S4 and no conserved charge-transfer centre in S2 (Fig. 3a, Extended Data Fig. 4a, b), and is therefore not predicted to form a functional voltage sensor. These predictions are corroborated by site-directed mutagenesis and electrophysiology¹⁰. Intervening lateral amphipathic helices on the cytoplasmic leaflet S4–S5 and S10–S11 link VSD1 and 2 to the pore domains P1 and P2, respectively, making polar and hydrophobic contacts with the pore gates 1 and 2, and providing connections for coupling voltage sensing, phosphorylation and binding of ions, inhibitors or proteins to the pore. S4–S5 is critical for coupling voltage to the pore gates⁹ in human TPC1.

Structural divergence between VSD1 and VSD2 generates the rectangular shape of TPC1. Variations in the S4 angle are found in symmetrical tetrameric ion channels, though the functional relevance is not clear (Extended Data Fig. 5). In the structure of Ca_v1.1, asymmetry of the pore domains P1–4 has a role in ion permeation and selectivity, whereas the VSDs remain close to symmetrical²⁰. TPCs and other tandem channels may use structural asymmetry to expand the repertoire of sensory responses and facilitate channel gating, as in the regulation of Ca_v channels by ions, toxins, pharmacophores, phosphorylation or additional subunits²¹.

Luminal Ca²⁺ inhibits plant TPC1 conductance^{10,11}. The plant TPC1 mutant D454N (also known as *fou2*) in the S7–S8 loop abolishes luminal Ca²⁺ sensitivity and increases the rate of channel opening¹¹. We observed two Ca²⁺-binding sites on the luminal side of VSD2.

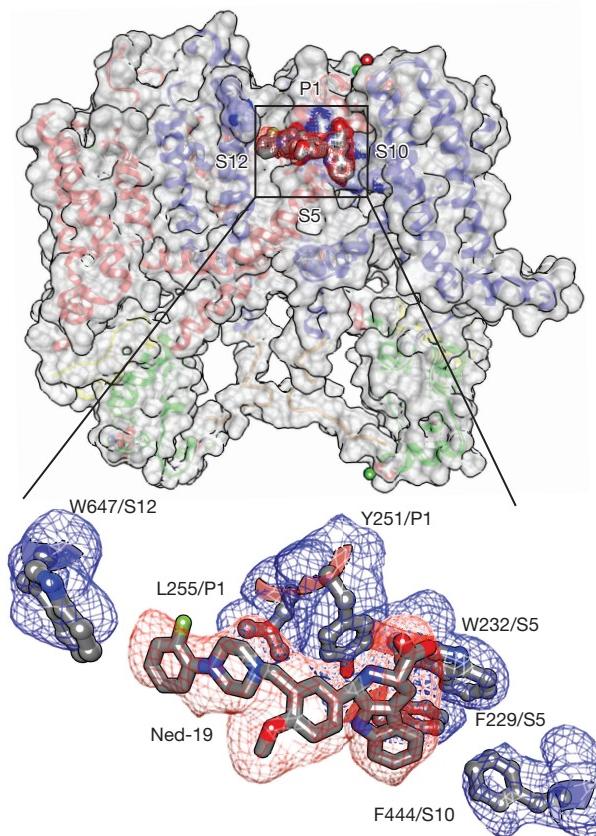


Figure 4 | Binding site for *trans*-Ned-19. Structural interactions between the TPC1 and Ned-19. Top, surface rendering showing the location of the Ned-19 binding site. Bottom, molecular details of the Ned-19 binding pocket at the interface between P1, P2 and VSD2. Simulated annealing omit density contoured at 1σ is shown for Ned-19 (red) and interacting side chains (blue).

Site 1 is coordinated by E457 from S7–S8 loop and E239 of P1, whereas site 2 uses D454 and D240 of P1 and E528 of S10. On the basis of mutagenesis and patch-clamp electrophysiology of TPC1, only site 2 is critical for luminal inhibition by Ca^{2+} ions¹⁰. Heavy-atom mimetics of Ca^{2+} (Ba^{2+} and Yb^{3+}) validate the placement of Ca^{2+} in sites 1 and 2,

defined by isomorphous and anomalous diffraction difference maps (Fig. 3, Extended Data Fig. 2c). A single fully occupied Yb^{3+} ion ($\text{Yb}2$) and Ba^{2+} site ($\text{Ba}1$) overlaps with site 1. In hTPC2, E528 is an aspartate, suggesting that there may be mechanistic similarities in ion and pH sensitivities between hTPC2 and plant TPC1. VSD1 binds a structural Ca^{2+} ion, confirmed by replacement with a Yb^{3+} ion ($\text{Yb}1$) coordinated by acidic residues E124 in S2, and D166 and D170 in S3 (Fig. 3a). Human and plant TPCs have an acidic residue E124 in S2, suggesting that this region could be a conserved Ca^{2+} -binding site. However, this ion-binding site has not been studied functionally.

Ned-19 (refs 7, 17) and Ca_v channel agonists and antagonists inhibit opening of TPC channels. Mutagenesis in human TPC1 demonstrated that L273 is important for channel activation and Ca^{2+} release²². The equivalent residues in TPC1 (M258 and L623) line P1 and P2 and form a hydrophobic surface that can bind hydrophobic ligands or stabilize the closed state in the absence of agonist (Extended Data Fig. 6). We observed density in TPC1 corresponding to the polycyclic TPC inhibitor Ned-19 (Fig. 4). Interactions with Ned-19 involve F229 in S5, both hydrophobic pi-stacking and a hydrogen bond between the carboxylate of Ned-19 and the side chain amino group of W232 in S5, L255 in P1, F444 in S10, and W647 in S12. Ned-19 effectively clamps the pore domains and VSD2 together, allosterically blocking channel activation. The binding site is consistent with Ned-19 as a high-affinity TPC-specific inhibitor^{7,17}. TPC1 S6 and S12 are 25–50% conserved with S6 helices in Ca_v channels (Extended Data Fig. 4c). Dihydropyridine and phenylalkylamine pharmacophores bind to specific aromatic residues at a nearby, yet distinct, region in S6 of the third and fourth tandem pore-forming domains in Ca_v channels²³. Homologous dihydropyridine-binding residues Y288, Y656, L663 and L664 in TPC1 and conserved residues in hTPC1 and hTPC2 could bind Ca_v pharmacophores in a parallel mechanism of inhibition⁷.

Ion selectivity varies among TPC channels. Plant TPCs are non-selective^{3,10}, whereas human TPC1 is Na^+ -selective^{5,9} and TPC2 conducts Na^+ , Ca^{2+} and possibly H^+ ions^{5,9,24}. Although the selectivity filter in TPCs is fairly conserved with Na_v and Ca_v channels, the specific anionic residues known to impart Na^+ or Ca^{2+} selectivity^{16,25} are replaced by hydrophobic side chains (Extended Data Fig. 4d). The largely hydrophobic selectivity filter of TPC1 agrees with the reported lack of ion selectivity¹⁰. A single anion in the selectivity filter of hTPC1 pore loop 2 E642 could confer Na^+ -selectivity as this residue, N624, faces the pore lumen in TPC1. Pore loop 1 forms a re-entrant

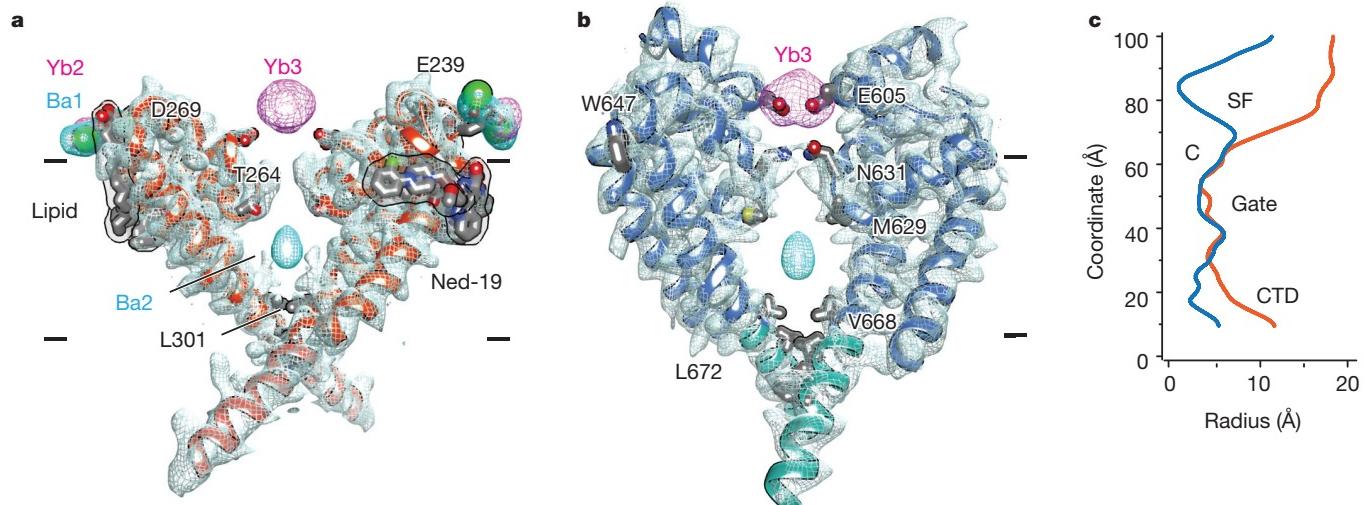


Figure 5 | The TPC1 ion channel. Cross-sectional views of the TPC1 ion channel separately through P1 and P2. Domains removed for clarity. **a, b**, P1 to P1 (orange; **a**) and P2 to P2 (blue; **b**) with sharpened $2mF_0 - DF_c$ electron density contoured at 1σ , and Ba^{2+} (cyan) and Yb^{3+} (magenta) isomorphous difference density contoured at 10σ . Membrane

boundaries are indicated by dashes. **c**, Pore radius calculations through separate pore domains, P1 (orange) and P2 (blue) using HOLE (see Methods). Approximate boundaries for the putative selectivity filter (SF), cavity (C), gates 1 and 2, and CTD are shown. Channel axis is vertical.

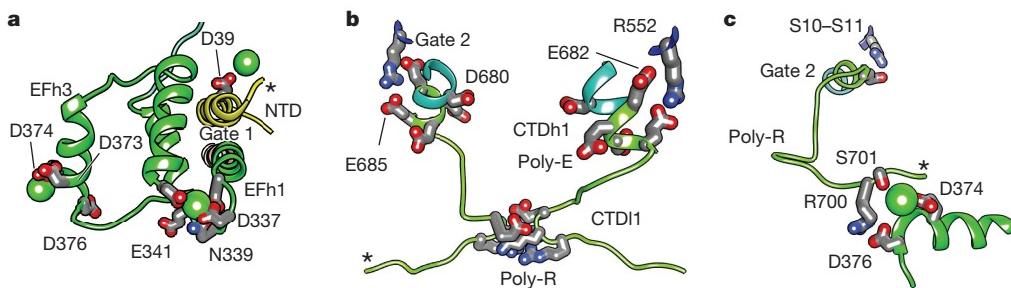


Figure 6 | Cytosolic sensory domains. Structure of EF-hand and coupling with NTD, VSD2 and CTD. **a**, Side views through the EF-hand (green) and NTD (yellow). **b**, Constriction site formed by gate 2 and CTD (chartreuse).

helix-turn-helix motif that forms a wide lining of two sides of the pore mouth (Fig. 5a). The high resolution data and crystallization conditions enabled visualization of electron density for lipid molecules, one bound to the backside of P2 at the interface of VSD2, flanking the luminal Ca^{2+} -binding site 2. Modelled as a palmitic acid, the lipid is a hydrogen-bond acceptor from T241 and M237 of P1 and bound adjacent to D240, the site for luminal Ca^{2+} inhibition. This could provide a basis for the reported modulation of TPCs by fatty acids (Supplementary Discussion). By contrast, pore loop 2 forms an extensive asymmetric constriction site at the pore mouth that brings four negative charges (D605 and D606) to its lining, of which D605 likely coordinates a Ca^{2+} ion (Fig. 5b), confirmed by Yb^{3+} ($\text{Yb}3$) observed in isomorphous and anomalous diffraction difference maps. The coordination distance of 2.5 Å suggests that this site could recognize ions from the lumen. However, this site is unlikely to function in selecting ions as the wide pore mouth, ~12–15 Å, of P1 does not provide a barrier to ion flow (Fig. 5c). D269 of P1 faces the pore, but with the distance to the lumen at 12.6 Å, it is too far away to function in ion selectivity.

Two N631 residues converge to 4.9 Å to form a narrow region that separates the upper vestibule from the putative selectivity filter. However, Q633 residues opposing P1 are 12.2 Å apart, leaving an opening for ion flow. This is the narrowest part of the selectivity filter.

Two T264 residues, conserved in Ca_v , Na_v and K_v subunits, define the upper boundary of the central cavity in TPC1 (Extended Data Fig. 4d, Fig. 5c). The central cavity contains solvent molecules and barium ions ($\text{Ba}2$), defined by isomorphous and anomalous difference maps. However, no ions are observed in the selectivity filter, unlike highly selective Na_v , Ca_v or K_v channels^{15,16,20}. The central cavity is filled with water and lined by hydrophobic residues. Two rings of hydrophobic residues L301, V668, Y305 and L672 seal the cavity towards the cytoplasm, and are part of the extensively studied pore gates of K_v channels^{26–28} and gates 1 and 2 in TPC1. On the cytosolic face of the gates, a K309–E673 salt bridge connects gate 1 and 2 together. K^+ ions have been observed in the gate and lower selectivity filter of the MlotiK1²⁹ and KcsA³⁰ potassium channels, and mutation of similarly positioned hydrophobic residues in the gate can modulate the channel state or flux of ions through the channel. This supports the idea that the VSDs, selectivity filter and gate undergo coordinated movements during channel opening and closing to regulate ion conductance.

In plant TPC1, two EF-hand domains between domains 1 and 2 confer sensitivity to cytosolic Ca^{2+} ions¹⁴. These domains consist of two Ca^{2+} -binding motifs EFL1 and EFL3, with 25–30% sequence conservation to human calmodulin (hCaM, Extended Data Fig. 4e). EFh1 forms a 70 Å long helix continuous with the gate helix S6 (Fig. 2b, Fig. 6a) that is connected to the first Ca^{2+} -binding site in EFL1 formed by side chain interactions with D335 (2.6 Å), D337 (2.8 Å), N339 (3.2 Å) and the backbone of E341 (2.3 Å). This site is probably structural, as it does not function in Ca^{2+} activation^{10,14}. The second site in EFL3 has been shown to facilitate channel activation by cytosolic Ca^{2+} , as a D376A mutant abolished all Ca^{2+} -dependent activation¹⁴. We observed

c, Interactions between EF-hand and CTD. Asterisks indicate phosphorylation sites. Transmembrane and adjacent domains removed for clarity.

a Ca^{2+} ion in EFL3 coordinated by E374 (2.9 Å) and D377 (5.2 Å), but surprisingly not by D376, which is 8.5 Å away. Thus the mechanism of Ca^{2+} activation may involve a structural change in EFL3, involving the Ca^{2+} -binding site and D376.

The CTD has an important role in plant TPCs (Supplementary Discussion), although it is least conserved across species (Extended Data Fig. 1). CTD helix 1 (CTDh1) continues from gate 2 and S12 with a poly-anionic (poly-E) helix E682–E685 (Fig. 6b). Ca_v channels contain a similar conserved poly-anionic region of variable length (Extended Data Fig. 4c). Poly-E residues E682 and E685 hydrogen-bond to R552 of S10–S11, directly linking the conformation of the CTDh1 to the voltage sensor VSD2 S10–S11. CTD loop 1 (CTDI1) from each monomer converges to form a charged constriction site, D694, K695 and R696 (Figs 5c and 6b), that brings six charged side chains to stabilize the channel through intermonomer contacts. The intermolecular constriction site partially includes a poly-R motif formed by R696, R698, R699 and R700, and may play a structural role or regulate binding of lipid second messengers or proteins (Extended Data Fig. 7). Notably, the CTD forms an intramolecular complex with the Ca^{2+} -activation site EFL3 (Fig. 6c). Surprisingly, R700 from CTDI1 makes a salt bridge with D376, the critical residue for Ca^{2+} activation. S701 coordinates the bound Ca^{2+} (4.2 Å) in EFL3, along with E374 and D377. These interactions include the potential site of phosphorylation S706 in CTDI1, not observed in the structure. Thus, reversible phosphorylation in the CTD could disrupt the EFh3–CTDI1 interaction and modulate Ca^{2+} sensing or couple directly to the VSD2 via the R552–E682–E685 interaction (Fig. 6c). The interactions between the regulatory sites and the central channel are outlined in Extended Data Fig. 7. The channel conductance of human and plant TPCs is multi-modal³. Our structure suggests a mechanism for channel opening, whereby Ca^{2+} concentrations, voltage and phosphoregulation are integrated through conformational changes in the VSDs, selectivity filter, gate, NTDs and CTDs to govern the conduction of ions (Extended Data Fig. 7).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 September 2015; accepted 2 February 2016.

- Hedrich, R. & Neher, E. Cytoplasmic calcium regulates voltage-dependent ion channels in plant vacuoles. *Nature* **329**, 833–836 (1987).
- Calcraft, P. J. et al. NAADP mobilizes calcium from acidic organelles through two-pore channels. *Nature* **459**, 596–600 (2009).
- Hedrich, R. & Marten, I. TPC1 – SV Channels Gain Shape. *Mol. Plant* **4**, 428–441 (2011).
- Patel, S. Function and dysfunction of two-pore channels. *Sci. Signal.* **8**, re7 (2015).
- Wang, X. et al. TPC proteins are phosphoinositide-activated sodium-selective ion channels in endosomes and lysosomes. *Cell* **151**, 372–383 (2012).
- Pitt, S. J. et al. TPC2 is a novel NAADP-sensitive Ca^{2+} release channel, operating as a dual sensor of luminal pH and Ca^{2+} . *J. Biol. Chem.* **285**, 35039–35046 (2010).
- Sakurai, Y. et al. Two-pore channels control Ebola virus host cell entry and are drug targets for disease treatment. *Science* **347**, 995–998 (2015).
- Hedrich, R., Flügge, U. I. & Fernandez, J. M. Patch-clamp studies of ion transport in isolated plant vacuoles. *FEBS Lett.* **204**, 228–232 (1986).

9. Cang, C., Bekele, B. & Ren, D. The voltage-gated sodium channel TPC1 confers endolysosomal excitability. *Nature Chem. Biol.* **10**, 463–469 (2014).
10. Guo, J. et al. Structure of the voltage-gated two-pore channel TPC1 from *Arabidopsis thaliana*. *Nature* <http://dx.doi.org/10.1038/nature16446> (this issue).
11. Beyhl, D. et al. The *fou2* mutation in the major vacuolar cation channel TPC1 confers tolerance to inhibitory luminal calcium. *Plant J.* **58**, 715–723 (2009).
12. Cang, C. et al. mTOR regulates lysosomal ATP-sensitive two-pore Na⁺ channels to adapt to metabolic state. *Cell* **152**, 778–790 (2013).
13. Morgan, A. J., Davis, L. C., Ruas, M. & Gallone, A. TPC: the NAADP discovery channel? *Biochem. Soc. Trans.* **43**, 384–389 (2015).
14. Schulze, C., Sticht, H., Meyerhoff, P. & Dietrich, P. Differential contribution of EF-hands to the Ca²⁺-dependent activation in the plant two-pore channel TPC1. *Plant J.* **68**, 424–432 (2011).
15. Long, S. B., Campbell, E. B. & Mackinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* **309**, 897–903 (2005).
16. Payandeh, J., Scheuer, T., Zheng, N. & Catterall, W. A. The crystal structure of a voltage-gated sodium channel. *Nature* **475**, 353–358 (2011).
17. Naylor, E. et al. Identification of a chemical probe for NAADP by virtual screening. *Nature Chem. Biol.* **5**, 220–226 (2009).
18. Kamb, A., Iverson, L. E. & Tanouye, M. A. Molecular characterization of Shaker, a *Drosophila* gene that encodes a potassium channel. *Cell* **50**, 405–413 (1987).
19. Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. Cloning of genomic and complementary DNA from Shaker, a putative potassium channel gene from *Drosophila*. *Science* **237**, 749–753 (1987).
20. Wu, J. et al. Structure of the voltage-gated calcium channel Cav1.1 complex. *Science* **350**, aad2395 (2015).
21. Catterall, W. A., Perez-Reyes, E., Snutch, T. P. & Striessnig, J. International Union of Pharmacology. XLVII. Nomenclature and structure-function relationships of voltage-gated calcium channels. *Pharmacol. Rev.* **57**, 411–425 (2005).
22. Brailoiu, E. et al. Essential requirement for two-pore channel 1 in NAADP-mediated calcium signaling. *J. Cell Biol.* **186**, 201–209 (2009).
23. Peterson, B. Z., Tanada, T. N. & Catterall, W. A. Molecular determinants of high affinity dihydropyridine binding in L-type calcium channels. *J. Biol. Chem.* **271**, 5293–5296 (1996).
24. Pitt, S. J., Lam, A. K. M., Rieddorf, K., Gallone, A. & Sitsapesan, R. Reconstituted human TPC1 is a proton-permeable ion channel and is activated by NAADP or Ca²⁺. *Sci. Signal.* **7**, ra46 (2014).
25. Tang, L. et al. Structural basis for Ca²⁺ selectivity of a voltage-gated calcium channel. *Nature* **505**, 56–61 (2014).
26. Armstrong, C. M. Time course of TEA⁺-induced anomalous rectification in squid giant axons. *J. Gen. Physiol.* **50**, 491–503 (1966).
27. Holmgren, M., Smith, P. L. & Yellen, G. Trapping of organic blockers by closing of voltage-dependent K⁺ channels: evidence for a trap door mechanism of activation gating. *J. Gen. Physiol.* **109**, 527–535 (1997).
28. Perozo, E., Cortes, D. M. & Cuello, L. G. Three-dimensional architecture and gating mechanism of a K⁺ channel studied by EPR spectroscopy. *Nature Struct. Mol. Biol.* **5**, 459–469 (1998).
29. Clayton, G. M., Altieri, S., Heginbotham, L., Unger, V. M. & Morais-Cabral, J. H. Structure of the transmembrane regions of a bacterial cyclic nucleotide-regulated channel. *Proc. Natl. Acad. Sci. USA* **105**, 1511–1515 (2008).
30. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K⁺ channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to J. M. Holton and G. Meigs at ALS beamline 8.3.1 and A. Gonzalez at SSRL beamline 12-2 for help with data collection and processing, B. P. Pedersen, B. H. Schmidt and J. S. Finer-Moore for critical analysis of the manuscript. This work was supported by NIH grant GM24485 to R.M.S. Beamline 8.3.1 at the Advanced Light Source is operated by the University of California Office of the President, Multicampus Research Programs and Initiatives grant MR-15-328599 and Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation. Figures were made with Chimera developed by the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311), and PyMOL developed by W. DeLano. Mass spectrometry was carried out by the UCSF mass spectrometry facility (supported by NIGMS P41-RR001614). The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract No. DE-AC02-05CH11231. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

Author Contributions A.F.K. performed the experiments, collected and processed the data, determined, refined, and analysed the structure. R.M.S. supervised the project and aided in structure determination and analysis. A.F.K. and R.M.S. wrote the manuscript.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under the accession number 5DQQ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M.S. (stroud@msg.ucsf.edu).

METHODS

No statistical methods were used to predetermine sample size.

Protein Production. *Arabidopsis thaliana* (At)TPC1 was expressed in codon-optimized form in *Saccharomyces cerevisiae*, using a plasmid based on p423_GAL1³¹ (Uniprot accession number Q94KI8). A C-terminal thrombin cleavage site was followed by a 12-residue glycine–serine linker and a 10-residue histidine tag. Transformed *S. cerevisiae* (strain DSY-5) were grown from overnight starter cultures in synthetic complete defined medium lacking histidine (CSM-HIS) in 1 l shaker flasks at 30 °C to OD_{600 nm} of 12–15. Protein expression was induced by adding 8% galactose dissolved in 4 × YP medium, to a final concentration of 2% galactose, at 30 °C for 18–20 h. Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris, pH 7.4, 500 mM NaCl, 1 mM EDTA, 1 mM phenylmethanesulfonyl fluoride (PMSF)), followed by bead beating using 0.5 mM beads for 30 min at 1 min intervals. The lysate was then centrifuged at 21,000g for 20 min, followed by collection of membranes by centrifugation at 186,010g for 150 min. Membranes were resuspended (50 mM Tris, pH 7.4, 500 mM NaCl, 10% glycerol), aliquoted, and frozen in liquid nitrogen for storage at –80 °C. Membrane aliquots of 7–12 g were solubilized in membrane buffer containing 1% (0.2 g per g of membrane) *n*-dodecyl-β-D-maltopyranoside (DDM), 0.1 mM cholesterol hemisuccinate (CHS), 0.03 mg ml^{−1} soy polar lipids, 1 mM CaCl₂, and 1 mM PMSF, for 1.5 h at 4 °C. The solubilized material was collected by centrifugation at 104,630g for 20 min, filtered through a 5 μm filter, and incubated with 3 ml of NiNTA beads in the presence of 20 mM imidazole, pH 7.4, for 7.5 h at 4 °C. NiNTA beads were collected by draining the mixture through a column and the beads were washed successively with 25 ml of NiNTA wash buffer (50 mM Tris, pH 7.4, 500 mM NaCl, 5% glycerol, 0.1% DDM, 0.1 mM CHS, 0.03 mg ml^{−1} soy polar lipids, 1 mM CaCl₂) plus 20 mM imidazole, pH 7.4, for 20 min at 4 °C, then 25 ml of NiNTA wash buffer plus 75 mM imidazole, pH 7.4, for 10 min at 4 °C, followed by 10 ml of NiNTA elution buffer (50 mM Tris, pH 7.4, 200 mM NaCl, 5% glycerol, 0.1% DDM, 0.1 mM CHS, 0.03 mg ml^{−1} soy polar lipids, 2 mM CaCl₂). TPC1 protein was obtained by incubating NiNTA beads with 200 units of bovine thrombin in 25 ml of NiNTA elution buffer overnight at 4 °C. The following day, NiNTA beads were washed with an additional 25 ml of NiNTA wash buffer plus 20 mM imidazole, pH 7.4, for 20 min at 4 °C, concentrated, and loaded on a Superose 6 column equilibrated in size-exclusion buffer (20 mM HEPES, pH 7.3, 200 mM NaCl, 5% glycerol, 0.05% DDM, 0.1 mM CHS, 0.03 mg ml^{−1} soy polar lipids, and 1 mM CaCl₂). Peaks fractions were pooled and concentrated to 5 mg ml^{−1} for crystallization. TPC1 and mutants were prepared similarly, except for the S701C mutant where 1 mM and 100 μM TCEP were included during solubilization and final size-exclusion chromatography, respectively.

Crystallization. Several eukaryotic TPCs, including hTPC1, hTPC2 and AtTPC1 were evaluated for expression in *S. cerevisiae* and SF9 cells, and homogeneity by fluorescence-based and conventional size-exclusion chromatography. TPC1 was selected based on expression level, stability, homogeneity and crystallization trials. Nevertheless, detergent-solubilized TPC1 was unstable and resistant to crystallization. Thermostability screening³² identified CHS, lipids and Ca²⁺ ions as improving TPC1 stability and homogeneity.

Wild-type TPC1 crystals were obtained in the presence of CaCl₂, CHS and soy polar lipids. Initial crystals from 200 nl drops in sparse matrix trays diffracted to ~10 Å resolution, with few diffracting to 4 Å. Reproducibility in 2 μl drops was achieved through extensive screening trials. To improve diffraction, truncation constructs were made. A construct lacking the N-terminal residues 2–11 (TPC1) diffracted on average to 7 Å, with 1% showing diffraction to 3.9 Å resolution, but were too sensitive to radiation and anisotropic to collect complete data sets. The TPC inhibitor *trans*-Ned-19 (Ned-19) was added to 1 mM final, 1 h before crystallization. These diffracted to ~5 Å resolution, with 50% to 3.9 Å resolution. Adding 100 μl of protein and 1 mg of DDM to a thin-film of 0.03 mg soy polar lipids with stirring overnight at 4 °C led to the TPC1 high lipid-detergent (HiLiDe)³³ mixture following ultracentrifugation for 20 min in a TLA-55 rotor at 112,000g. Average diffraction was 4 Å, with 10% diffracting to 3.5–3.7 Å resolution, the highest resolution being 3.5 × 5.5 × 4.0 Å. Anisotropic resolution was determined using the CCP4 program Truncate³⁴ and the UCLA Anisotropy Server³⁵. Large improvements were achieved through partial dehydration. HiLiDe crystals were incubated with additional amounts of polyethylene glycol 300, 0–20% (v/v), before they were collected. Incubating with additional 15% PEG300, increased the average diffraction to 4 Å, best diffraction to 2.7 Å resolution, and diminished the diffraction anisotropy. Derivative crystals were obtained by soaking in 0.1–50 mM of heavy-atom compounds in a 1:1 mixture of crystal buffer (0.1 M glycine pH 9.3, 50 mM KCl, 1 mM CaCl₂, 35% PEG300) for 24–48 h at 20 °C. Partial dehydration was performed after derivatization.

Data collection and reduction. X-ray diffraction was collected at the Advanced Light Source Beamlines (ALS) 8.3.1, 5.0.2 and the Stanford Synchrotron Radiation Laboratory Beamline (SSRL) 12-2. Data were reduced using XDS³⁶. Native data

were collected at wavelength 1.000 Å. Data from heavy-atom derivative crystals were collected at the anomalous peak of the L-III edge for each element (Hg = 1.006 Å, Ta = 1.2546 Å, Yb = 1.3854 Å), except for Ba, which is collected at 1.750 Å.

Structure determination of TPC1. *De novo* structure determination was based on nine heavy metal and cluster derivatives at 3.5 Å (Extended Data Table 1). Dehydration improved diffraction to 2.7 Å while diminishing the anisotropy (Extended Data Table 2 and 3). Experimental phases were calculated by multiple isomorphous replacement with anomalous scattering (MIRAS) from both non-dehydrated (Extended Data Table 1) and dehydrated (Extended Data Table 2) crystals derivatized with tantalum bromide clusters (Ta₆Br₁₂), BaCl₂, YbCl₃, and a mutant of TPC1 derivatized with HgCl₂.

A single tantalum atom was identified by single isomorphous replacement with anomalous scattering (SIRAS) using SHELXC/D³⁷. A Ta₆Br₁₂ cluster was refined and phases calculated using SHARP³⁸. These phases were used to place sites in other derivatives and calculate phases using SHARP³⁸. Initial MIRAS phases were determined with non-dehydrated crystals with 100 μM Ta₆Br₁₂ clusters (derivative 1; 1 site, isomorphous phasing power (PP_{iso}; defined as PP_{iso} = root mean square (r.m.s.) ($|F_H|_{\text{cal}}/\sum |F_{\text{PH}} - F_{\text{P}}|_{\text{obs}} - |F_{\text{PH}} - F_{\text{P}}|_{\text{calc}}$) or r.m.s. ($|F_H|_{\text{cal}}/\sum \varepsilon_{\text{iso}}$), where $|F_H|$ is the heavy-atom structure factor amplitude, $|F_{\text{PH}}$ the heavy-atom derivative structure factor amplitude, $|F_{\text{P}}$ the native structure factor amplitude, and ε_{iso} the phase-integrated lack of closure for isomorphous differences) of 1.0/0.77 (ancentric/centric); anomalous phasing power (PP_{ano}; defined as PP_{ano} = r.m.s. ($\sum |F_{\text{PH}+} - F_{\text{PH}-}|_{\text{cal}}/\sum \varepsilon_{\text{ano}}$) or r.m.s. ($\sum |F_{\text{PH}+} - F_{\text{PH}-}|_{\text{cal}}/\sum |F_{\text{PH}+} - F_{\text{PH}-}|_{\text{obs}} - |F_{\text{PH}+} - F_{\text{PH}-}|_{\text{calc}}$), where $|F_{\text{PH}+} - F_{\text{PH}-}$ is the mean anomalous difference between Friedel pairs, and ε_{ano} the phase-integrated lack of closure for anomalous differences) of 0.31; isomorphous Cullis R factor (R_{Cullis,iso}; defined as R_{Cullis,iso} = $\sum \varepsilon_{\text{iso}}/\sum |F_{\text{PH}} - F_{\text{P}}|_{\text{obs}}$) of 0.766/0.755 (ancentric/centric); anomalous Cullis R factor (R_{Cullis,ano}; defined as R_{Cullis,ano} = $\sum \varepsilon_{\text{ano}}/\sum |F_{\text{PH}+} - F_{\text{PH}-}|_{\text{obs}}$) of 0.584), 1 mM Ta₆Br₁₂ clusters (derivative 2; 1 site, PP_{iso} of 2.32/1.94 (ancentric/centric), PP_{ano} of 0.761, R_{Cullis,iso} of 0.525/0.536 (ancentric/centric), R_{Cullis,ano} of 0.792), or 2 mM Ta₆Br₁₂ clusters (derivative 3; 1 site, PP_{iso} of 0.89/1.1 (ancentric/centric), PP_{ano} of 1.36, R_{Cullis,iso} of 0.850/0.765 (ancentric/centric), R_{Cullis,ano} of 0.572), using merged data (native 1) from three non-dehydrated crystals as a source of isomorphous native amplitudes (Extended Data Table 1).

This map allowed building of all transmembrane segments and partial building of the EF-hand domain in TPC1. Side-chain features and density for loop regions were largely absent from calculated maps, precluding subunit identification, and objective determination of the helical register. Initial MIRAS phases were further improved by including merged data from five non-dehydrated crystals derivatized with 50 mM BaCl₂ (derivative 4; 3 sites, PP_{iso} of 0.75/0.58 (ancentric/centric), PP_{ano} of 0.59, R_{Cullis,iso} of 0.905/0.888 (ancentric/centric), R_{Cullis,ano} of 1.00) in the MIRAS calculation, resulting in an overall figure of merit (FOM) of 0.262/0.333 (ancentric/centric) (Extended Data Table 1). Combined tantalum and barium phases allowed fitting of some side chains to density, placement of ions binding sites in the channel lumen and VSDs (Extended Data Fig. 2, Supplementary Data Table 1). The source of phases with highest resolution came from derivatives obtained from crystals that were dehydrated after a 24–48 h soak in heavy-atom solution. MIRAS phases were calculated from native crystals derivatized with 1 mM Ta₆Br₁₂ clusters (derivative 5; 1 site, PP_{iso} of 4.09/3.56 (ancentric/centric), PP_{ano} of 2.42, R_{Cullis,iso} of 0.302/0.286 (ancentric/centric), R_{Cullis,ano} of 0.495); and derivative 6: 1 site, PP_{iso} of 8.54/8.91 (ancentric/centric), PP_{ano} of 2.29, R_{Cullis,iso} of 0.149/0.122 (ancentric/centric), R_{Cullis,ano} of 0.517) or 2 mM Ta₆Br₁₂ clusters (derivative 7: 1 sites PP_{iso} of 3.65/3.54 (ancentric/centric), PP_{ano} of 2.25, R_{Cullis,iso} of 0.372/0.323 (ancentric/centric), R_{Cullis,ano} of 0.535), 1 mM YbCl₃ (derivative 8: 4 sites, PP_{iso} of 0.851/0.795 (ancentric/centric), PP_{ano} of 0.699, R_{Cullis,iso} of 0.962/0.884 (ancentric/centric), R_{Cullis,ano} of 0.982), and a TPC1 S701C mutant derivatized with 1 mM HgCl₂ (derivative 9: 5 sites, PP_{iso} of 0.792/0.822 (ancentric/centric), PP_{ano} of 0.736, R_{Cullis,iso} of 0.935/0.704 (ancentric/centric), R_{Cullis,ano} of 1.02), followed by dehydration resulting in a FOM of 0.267/0.341 (ancentric/centric) (Extended Data Table 2). Non-isomorphism errors were minimized by using merged data from three non-dehydrated crystals extending to 3.5 Å resolution (native 1) as a source of isomorphous native amplitudes for all phasing calculations, density modification, and initial model building (Extended Data Tables 1 and 2). MIRAS phases from dehydrated derivatives extend to higher resolution, contain less diffraction anisotropy, and display much clearer electron density for main-chain and side-chain features, and allow objective determination of helical register (Extended Data Fig. 2b). Mercury heavy atom positions were used as restraints, along with sequence homology, for objective placement of cysteine residues (Supplementary Data Table 1). Ytterbium and barium heavy-atom positions were used in conjunction with homology for placing likely (D, E) chelate residues within the pore, VSDs, and pore (Extended Data Fig. 2c, Supplementary Data Table 1). MIRAS phases from non-dehydrated and dehydrated derivatives were further improved through solvent flipping in SOLOMON³⁹ with 75% and 70% solvent content and FOMs of 0.637

and 0.470. Phases were then combined and extended to 3.5 Å resolution through iterative cycles of density modification in DMMULTI⁴⁰, using histogram matching, solvent flattening, gradual phase extension⁴¹, and threefold cross-crystal averaging with un-phased, dehydrated, native amplitudes (native 2) with 70% solvent content resulting in a FOM of 0.493 (Extended Data Fig. 2b). Averaging masks were generated by NCSMASK³⁴ with radius of 4 Å from atomic models from each iteration of building and refinement. Model phases were purposely left out of experimental density modification procedures to reduce bias.

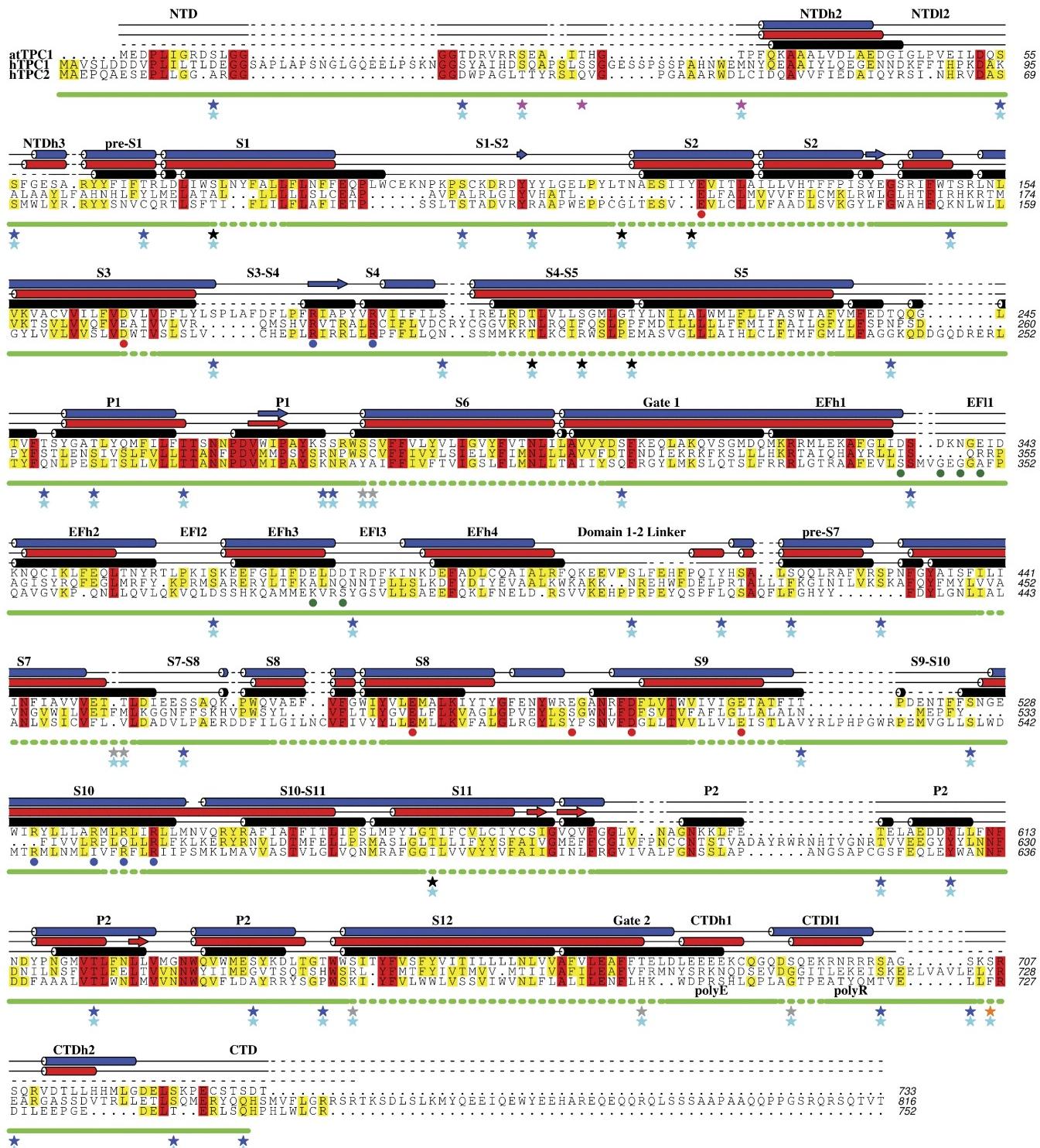
Initial model building was guided by the position of heavy atoms determined from isomorphous difference maps (Extended Data Fig. 2c, Supplementary Data Table 1). Using anomalous differences yielded equivalent maps and structural restraints.

Refinement. The best native data (native 2) extends to an overall resolution of $2.8 \times 4.0 \times 3.3$ Å (ref. 35), and enabled building of 85% of the TPC1 protein sequence. Structure interpretation was performed using COOT⁴², with refinement in PHENIX⁴³. Simulated-annealing was employed alongside refinement of coordinates, TLS restraints, and several cycles of automated building using ARP/wARP³⁴ in CCP4. Building of the NTD, CTD, and loop regions were guided by restraints based on heavy atom positions (Supplementary Data Table 1), omit maps, and gradual improvement of the map and crystallographic *R* factors. Refinement of the structure in triclinic space group P1 led to a substantial reduction in *R*_{free} (39.9%). During review of our manuscript, a report of the crystal structure of TPC1 was published¹⁰. The coordinates of AtTPC1 are available under PDBID accession number 5EIJ. The topology of the two structures were identical and their analysis had determined the registration of ~14 additional sites by mutation of residues to cysteine. Therefore, we checked our structure against that reported in ref. 10 for registration, leading to several corrections of registration in the sequence and improvements in map quality and crystallographic *R* factors (Extended Data Table 3). Anisotropy-corrected data ($2.8 \times 4.0 \times 3.3$ Å) were used during the final stages of refinement with an applied isotropic B-factor of -78.72 Å². Our structure of TPC1 was refined to 2.87 Å resolution with final *R*_{work}/*R*_{free} of 29.71% and 33.89%. Analysis by Molprobity shows Ramachandran geometries of 91.8%, 6.9% and 1.3% for favoured, allowed and outliers, respectively. The structure contains 17.0% rotamer outliers.

Domain assignment. On the basis of experimental electron density (Extended Data Fig. 2b), heavy-atom restraints (Extended Data Fig. 2c and Supplementary Data Table 1), and sequence homology (Extended Data Fig. 1), we have assigned the TPC1 domain structure as follows (Fig. 2a, Extended Data Fig. 1): N-terminal domain (NTD; residues 30–55), pre-S1 helix (pre-S1; residues 60–71), voltage-sensing domain 1 (VSD1; S1–S4, residues 72–215), pore domain 1 (P1; S5–S6, residues 216–301), gate 1 (residues 302–321), EF-hand (EF; residues 322–399), domain 1–2 linker (residues 400–418), pre-S7 helix (pre-S7; residues 419–433), voltage-sensing domain 2 (VSD2; S7–S10, residues 434–564), pore domain 2 (P2; S11–S12, residues 565–668), gate 2 (residues 669–685), C-terminal domain (CTD; residues 686–703). Unstructured regions include part of the NTD (residues 1–29, 56–59), S3–S4 (residues 174–183), domain 1–2 linker (residues 408–412), S9–S10 (residues 519–523), P2 (residues 591–593), and CTD (residues 704–733) for a total of 84 missing residues (11.5% unstructured). The structure of TPC1 comprises 649 residues (88.5% of total sequence). The restraints used for structure determination are summarized in Supplementary Data Table 1.

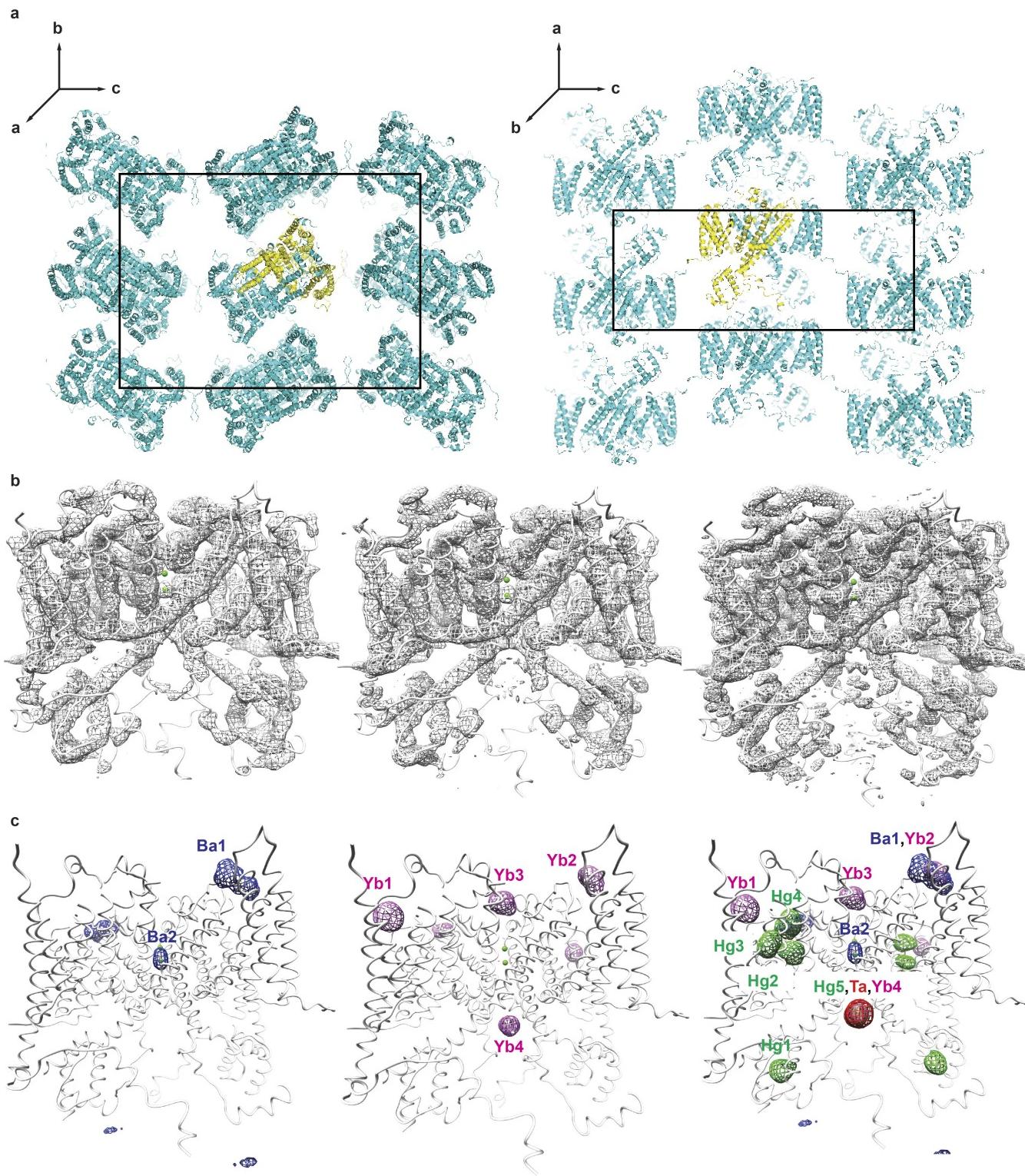
Sequence alignments were performed using MUSCLE⁴⁴ and ALINE⁴⁵. Structural figures were made using Chimera⁴⁶ and PyMOL. Pore radius calculations using HOLE⁴⁷.

31. Mumberg, D., Müller, R. & Funk, M. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22**, 5767–5768 (1994).
32. Hattori, M., Hibbs, R. E. & Gouaux, E. A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Structure* **20**, 1293–1299 (2012).
33. Gourdon, P. et al. HiLiDe—systematic approach to membrane protein crystallization in lipid and detergent. *Cryst. Growth Des.* **11**, 2098–2106 (2011).
34. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
35. Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
36. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
37. Sheldrick, G. M. Experimental phasing with SHELXC/D/E : combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
38. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
39. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
40. Cowtan, K. ‘dm’: an automated procedure for phase improvement by density modification, in Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography. 31, 34–38 http://www ccp4.ac.uk/newsletters/newsletter36/07_dm.html (1994).
41. Pedersen, B. P., Morth, J. P. & Nissen, P. Structure determination using poorly diffracting membrane-protein crystals: the H⁺-ATPase and Na⁺,K⁺-ATPase case history. *Acta Crystallogr. D* **66**, 309–313 (2010).
42. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
43. Adams, P. D. et al. PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
44. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
45. Bond, C. S. & Schüttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D* **65**, 510–512 (2009).
46. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
47. Smart, O. S., Neduvelli, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996).
48. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices1. *J. Mol. Biol.* **292**, 195–202 (1999).
49. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892–893 (1998).
50. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649 (2004).
51. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013).
52. Paulsen, C. E., Armache, J.-P., Gao, Y., Cheng, Y. & Julius, D. Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature* **520**, 511–517 (2015).



Extended Data Figure 1 | Sequence alignment of TPCs with TPC1 experimental and predicted secondary structure. A sequence alignment based on seven human and plant TPC orthologues with observed TPC1 (black), and secondary structure predictions by Psipred⁴⁸ (red) and Jpred⁴⁹ (blue). Helices are shown as cylinders, β sheets as planks, coils as solid lines, and unstructured regions as dashed lines. Level of conservation is indicated by colour (>50% yellow; >80% red). Blue dots mark arginines in S4; red dots mark charge-transfer anions; green dots mark Ca^{2+} -binding

sites in the EF-hand. Solid and dashed green lines represent observed and absent peptides from mass spectrometry experiments. Coloured stars indicate observed phosphorylation sites using electrospray ionization mass spectrometry (magenta; Extended Data Fig. 3a), potential phosphorylation sites identified by truncation constructs (orange; Extended Data Fig. 3e), predicted phosphorylation sites by NetPhosK⁵⁰ (cyan), non-phosphorylated sites (blue), unlikely sites owing to solvent inaccessibility (black), and unknown or unidentified regions (grey).



Extended Data Figure 2 | Crystal packing of TPC1 and experimental electron densities. **a**, Views of the TPC1 C222₁ crystal lattice viewed along two-fold axes parallel to *a* (left) and *b* (right). Unit cell boundaries are shown. The asymmetric unit is shown in yellow. **b**, Transverse view of TPC1 is shown with overlayed FOM-weighted experimental electron density calculated using native amplitudes (native 1) and heavy-atom phases contoured at 1 σ . Density-modified phases from non-dehydrated

derivatives (left; Extended Data Table 1), dehydrated derivatives (middle; Extended Data Table 2), and combined phases with solvent flattening, histogram matching, phase extension to high resolution native (native 2), and cross-crystal averaging (right; see Methods, Extended Data Table 3). **c**, Transverse views of TPC1 with overlayed heavy-atom electron densities calculated from isomorphous differences.

a Observed peptides

MEDPLIGRDSLGGGGTDRVRRSEAITHGTPFQKAAALVDLAEDGIGLPVEILDQSSFGES 60
 ARYYFIFTRLDLILWSLNYYFALLFLNFFEQLWCEKNPKPSCKDRDYYLGELPYLTNAES 120
 IIYEVITLAILLVHTFFFPISEGSRIFWTSRLNLVKVACCVVILFVDVLVD^{FLY}LSP^{LA}F 180
 FLPFRIAPYVRVIIFILSIR^ELRDTLVLLSGMLGTYLNILALWMLFLFASWIAFVMFED 240
 TQQGLTVFTSYGATLYQMFI^LFTTSNNPDVWI^PAYKSSR^WSSVFFVLYVLIGVYFVTNLI 300
 LAVVYDSFKEQLAKVQSGMDQM^KRRL^KLEKA^GFLIDSDKNGE^IDKNQC^IKLFEQLTNYRTL 360
 PKISKEEFG^LIFDELDDTRDFKINKDEFADIC^AIALRFQKEEVPSL^FEHFPQIYHSALS 420
 QQLRAFVRSPNFGYAISF^ILIINFI^VAVVVETTL^DIEESSA^QKPWQVAEFVFGWIYVLEMA 480
 LKIYTYGFENYWR^EGANRFDFL^TWVIVIGETATF^ITPDENTFFSN^GEWIRYLLLARMLR 540
 LIR^LLMNVQRYRAFIATFITLIPSLMPY^LG^TI^CVL^CIYCSIGVQVF^GGLVNAGNKKLFE 600
 TELAEDDYLLFNFDNYPNGMV^TLFNLLVMGNWQVWMESY^KDLTG^TWWSITYFVSFYVITI 660
 LLLLNLNVAFVLEAFFTEL^DLEEEEK^CQGQDSQEK^RNRRRSAGSK^SRQRVDTLLHHMLG 720
 DELSKPECSTS^DT 733

Prediction

1	*	*	*	*	*	*	**	60
MEDPLIGRDSLGGGGTDRVRRSEAITHGTPFQKAAALVDLAEDGIGLPVEILDQSSFGES								

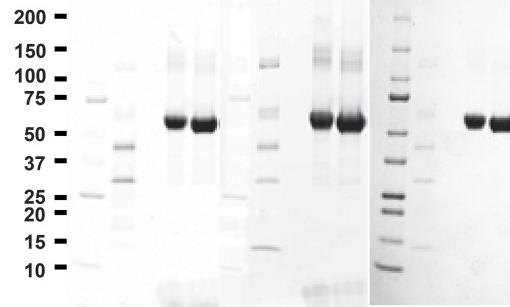
Observed phosphopeptides

GGGGTDRV^RSEA^I^THGT^PF^K
 VRR^SEA^I^THGT^PF^K
 RSE^AI^THGT^PF^K
 SEA^I^THGT^PF^K

b

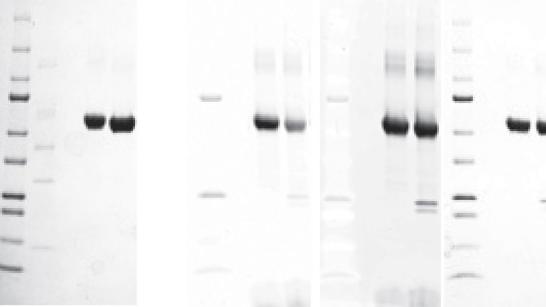
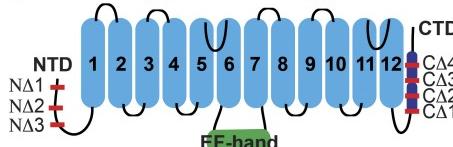
Phosphoprotein Sypro

WT TPC1

**c**

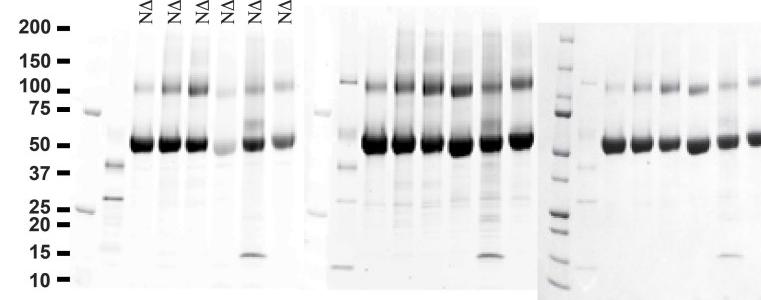
Phosphoprotein Sypro

WT TPC1

**d****e**

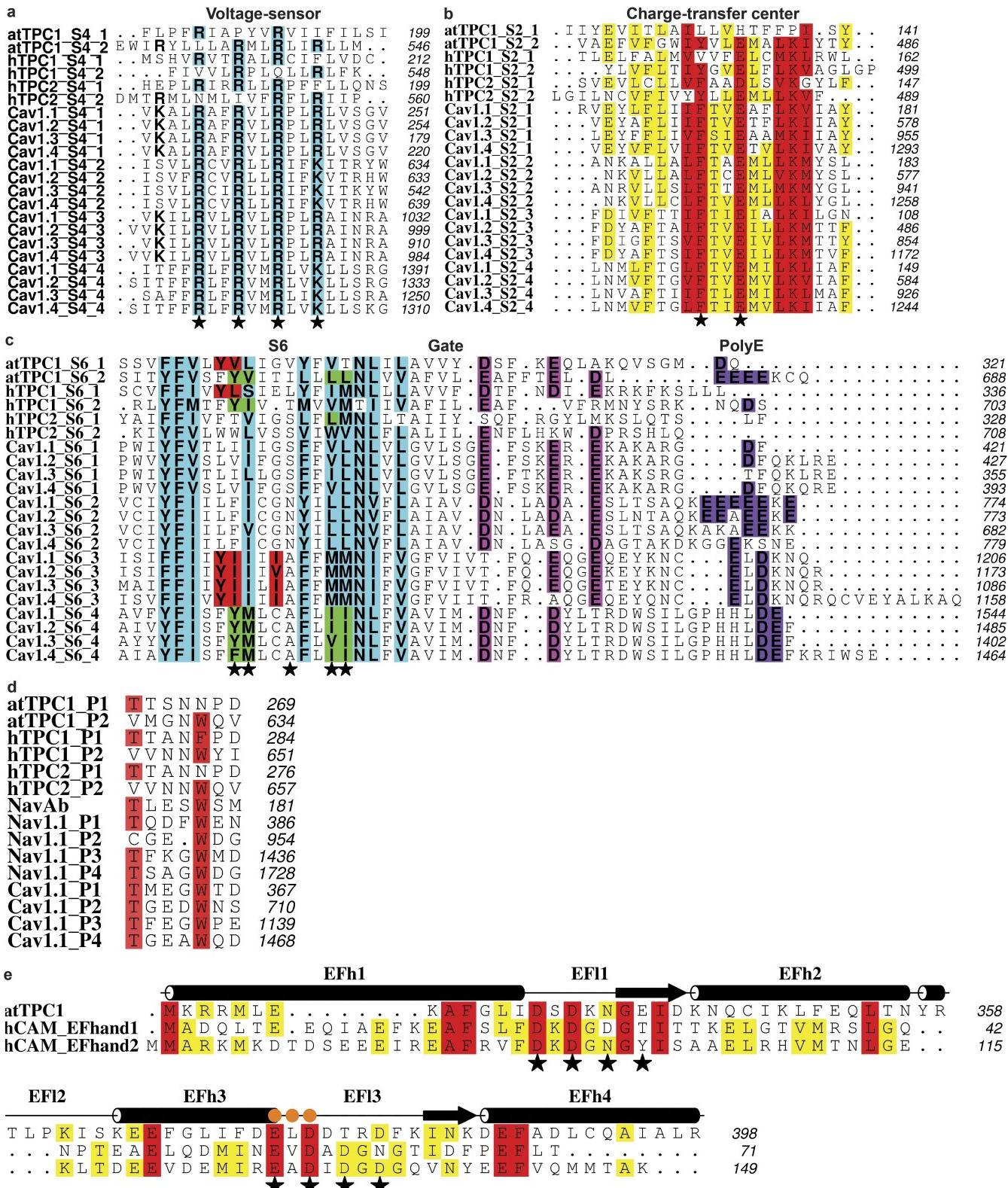
Phosphoprotein Sypro

Coomassie



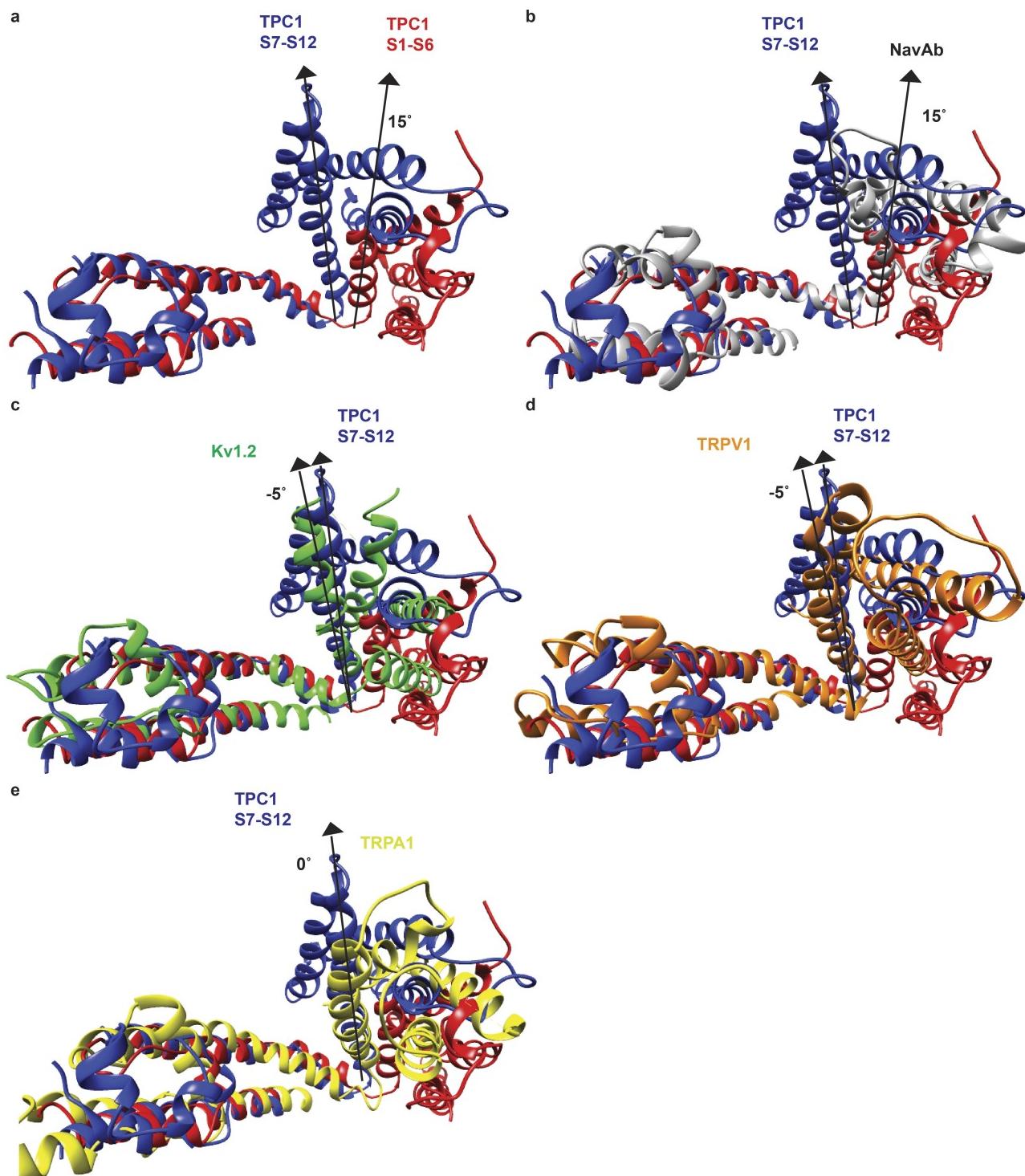
Extended Data Figure 3 | Determination of phosphorylation sites in TPC1. **a**, Electrospray ionization mass spectrometry peptide sequence coverage from an in-gel digest of wild-type (WT) TPC1 by four enzyme combinations (Trypsin/Asp-N, Trypsin/Glu-C, Lys-C or Chymotrypsin). Measured peptides (top, red highlight), predicted (middle, black star) and observed (middle, red star) phosphorylation sites are shown. Observed phosphopeptides are listed with the sites of phosphorylation coloured red. **b–e**, Polyacrylamide gels of purified TPC1 (10 μ g) stained with phosphoprotein-specific probe ProDiamond-Q, SyproRuby or Coomassie

(left to right). Molecular weights of standards are indicated in kilodaltons. The first two lanes are PrecisionPlus protein molecular weight standards and PeppermintStick phosphoprotein molecular weight standards. **b**, Wild-type TPC1 and crystallographic TPC1. **c**, Untreated and treated wild-type TPC1 with lambda phosphatase for 1 h at 25 °C. **d**, Schematic of TPC1 truncations (N Δ 1; 2–11, N Δ 2; 2–21, N Δ 3; 2–30, C Δ 1; 682–733, C Δ 2; 693–733, C Δ 3; 709–733, C Δ 4; 724–733). **e**, Analysis of TPC1 N- and C-terminal truncations for binding to ProDiamond-Q. N Δ 1C Δ 1 was unstable during purification.



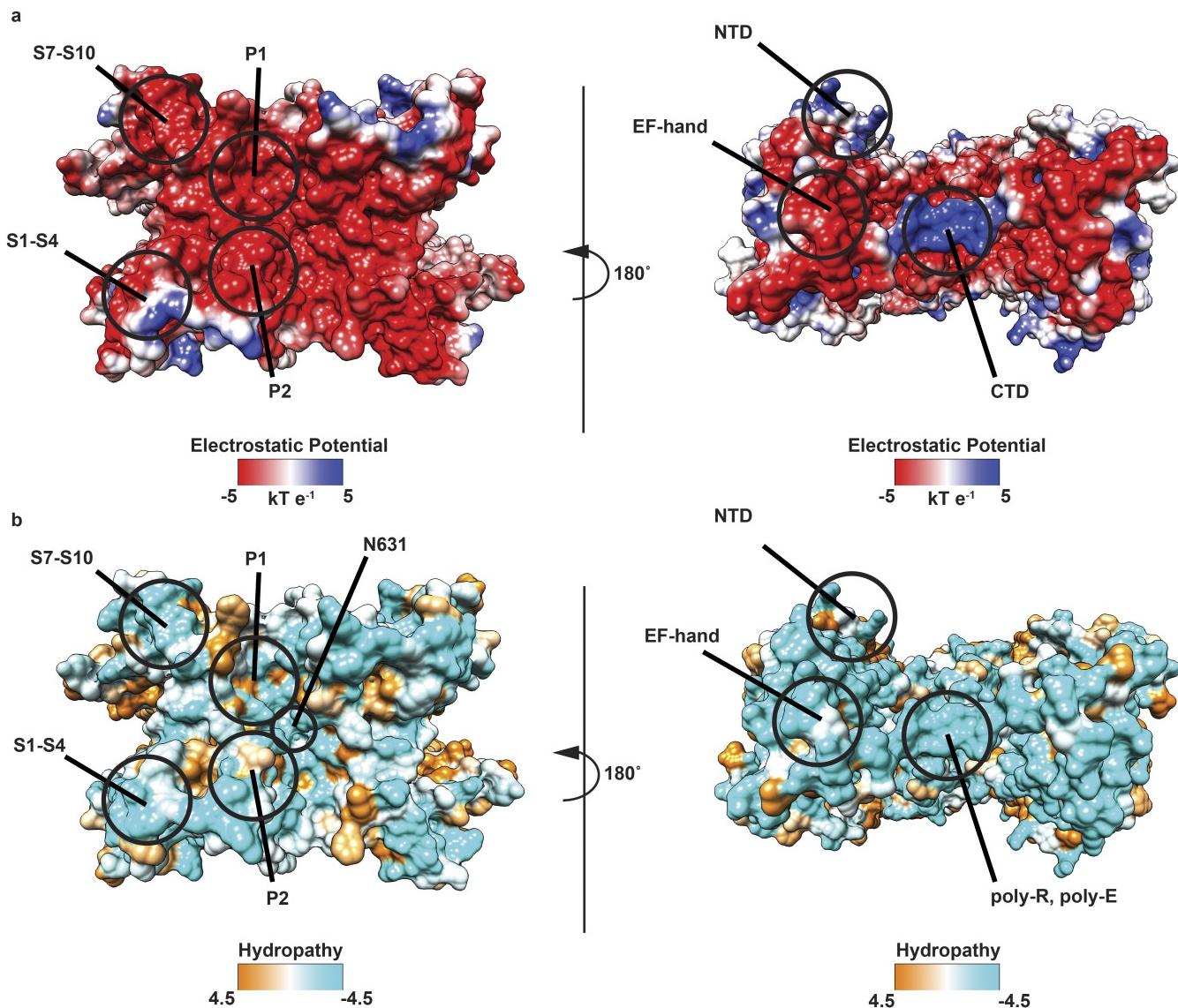
Extended Data Figure 4 | Sequence alignment of TPC1 subdomains with hTPC1, hTPC2, human Ca_v 1.1–1.4, Na_v 1.1, Na_vAb and hCaM. **a**, S4 voltage-sensing segments. Conserved arginines are highlighted in cyan. Stars mark potential voltage-sensing residues. **b**, S2 segments. Stars mark conserved charge-transfer residues. **c**, S6 segments, pore gate and poly-E motif. Conserved hydrophobic residues are highlighted in cyan. Dihydropyridine-binding residues in the Ca_v S6 domain 3, domain 4, and corresponding residues in TPC1 are highlighted in red and green,

respectively. Stars mark the position of phenylalkylamine drug-binding residues in Ca_v s. Conserved residues in the gate and poly-E motif are highlighted in magenta and purple, respectively. **d**, Pore loops. Conserved residues are highlighted in red. **e**, Alignment of hCaM EF-hand domains (EF-hand 1, residues 1–71; EF-hand 2, residues 72–149) and AtTPC1 (residues 322–398). Stars mark calcium binding motifs and orange dots mark the interaction site with CTD.



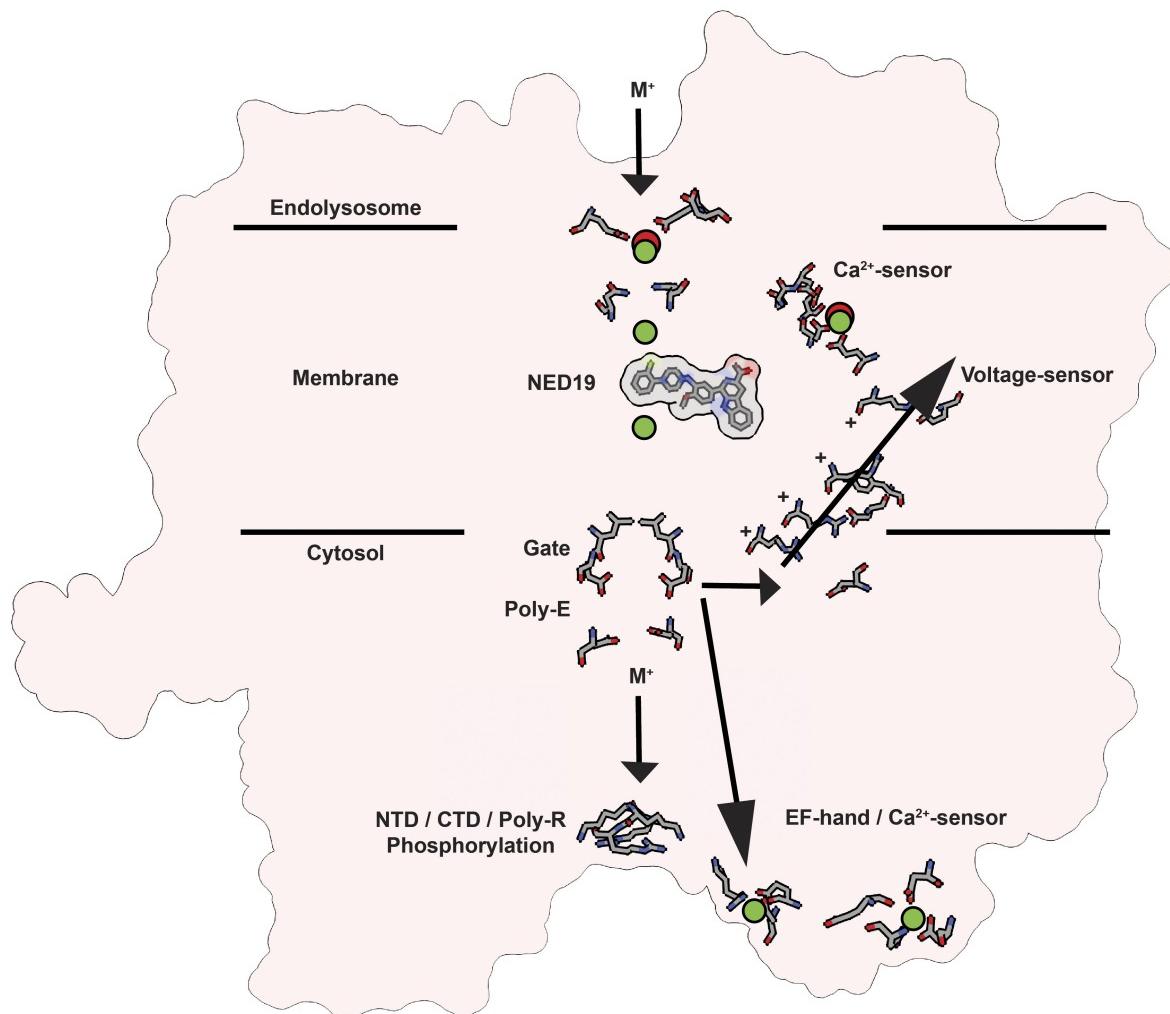
Extended Data Figure 5 | Comparison of VSDs between TPC1 and symmetrical ion channels. **a–e**, Structural alignment of S11–S12 segments of TPC1 domain 2 (blue) with S5–S6 of TPC1 domain 1 (**a**; red),

$\text{Na}_\text{v} \text{Ab}$ ¹⁶ (**b**; PDB ID 3RVY, grey); $\text{K}_\text{v} 1.2$ (**c**; PDB ID 2A79, green)¹⁵, TRPV1 (**d**; PDB ID 3J5P, orange)⁵¹, and TRPA1 (**e**; PDB ID 3J9P, yellow)⁵². Angles between S4 segments with respect to S10 of TPC1 domain 2 are shown.



Extended Data Figure 6 | Electrostatic surfaces and hydrophobic surfaces. **a, b,** Top from the luminal side (left) and bottom from the cytoplasmic side (right) views of an electrostatic surface representation (a) and surface representation, coloured according to Kyte–Doolittle hydrophobicity of TPC1 (b). Notable domains and residues are labelled.

Electrostatic potential and Kyte–Doolittle hydrophobicity without any bound ions were generated using Chimera⁴⁶. The EF-hand and NTD domains are negatively charged and bind cations. The poly-R motif accounts for the positively charged region in the CTD.



Extended Data Figure 7 | Mechanism for TPC gating. A schematic summarizing structural features of TPC1 that suggest mechanisms for voltage-sensing, ion permeation, inhibition by Ned-19, luminal Ca^{2+} inhibition, cytosolic Ca^{2+} -activation (EF-hand), and phosphoregulation

(NTD/CTD). Ca^{2+} (green) and lanthanide (red) ion binding sites are shown. An ion permeation pathway through the putative selectivity filter, gate, and poly-E and poly-R motifs are summarized. M^+ represents a general cation (Na^+ , K^+ , Ca^{2+}) and '+' signs are gating charges.

Extended Data Table 1 | Data collection and phasing statistics for non-dehydrated derivatives

	Native 1 (Isomorph to der. 1-9) ^a	Derivative 1 $\text{Ta}_6\text{Br}_{12}$	Derivative 2 $\text{Ta}_6\text{Br}_{12}$	Derivative 3 $\text{Ta}_6\text{Br}_{12}$	Derivative 4 BaCl_2
Data collection					
Space group	$\text{C}222_1$	$\text{C}222_1$	$\text{C}222_1$	$\text{C}222_1$	$\text{C}222_1$
Cell dimensions					
a, b, c (Å)	88.58, 161.39, 219.4	89.05, 161.0, 223.08	89.05, 161.0, 220.59	89.14, 164.87, 218.63	88.93, 166.20, 219.48
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.000	1.25465	1.25465	1.25465	1.750
Resolution (Å)	20-3.5(4- 3.5) ^b	20-4(5-4)	20-5(6-5)	20-4(5-4)	20-4(5-4)
R_{meas} (%) ^c	28.8(437.2)	22.3(181.2)	20.6(683.3)	45.3(899.5)	29.5(532.1)
$I/\sigma I$	7(0.84)	6.2(1.5)	4.5(0.34)	2.8(0.3)	6.97(0.74)
$\text{CC}_{1/2}$ ^d	99.9(72.5)	99.8(88.6)	99.6(23.8)	99.6(28.5)	99.9(76.2)
Completeness (%)	97.8(97.4)	99.1(99.7)	98.4(99.7)	98.5(98.9)	99.1 (99.7)
No. reflections	19817	25631	13926	25944	26349
Redundancy	7.6(7.8)	7.2(7.1)	7.2(7.3)	7.4(7.5)	31(23)
Anisotropy ^e	3.5 x 5.5 x (a, b, c)	4.0 x 6.0 x 4.0 Å	6.2 x 7.5 x 6.2 Å	4.5 x 6.3 x 6.8 Å	4.2 x 6.3 x 5.4 Å

^aNative 1, 3 crystals; derivative 1, 1 crystal; derivative 2, 1 crystal; derivative 3, 1 crystal; derivative 4, 5 crystals.^bHighest resolution shell is shown in parentheses.^cRedundancy of independent R factor calculated in XDS.^dPercentage of correlation between intensities from random half-datasets, as calculated in XDS.^eCalculated anisotropy from UCLA diffraction anisotropy server.

Extended Data Table 2 | Data collection and phasing statistics for partially dehydrated derivatives

	Derivative 5 Ta₆Br₁₂^a	Derivative 6 Ta₆Br₁₂	Derivative 7 Ta₆Br₁₂	Derivative 8 YbCl₃	Derivative 9 HgCl₂
Data collection					
Space group	C222 ₁	C222 ₁	C222 ₁	C222 ₁	C222 ₁
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	88.57, 156.41, 218.71	88.97, 159.68, 218.71	89.32, 159.58, 219.37	89.29, 162.23, 220.93	89.47, 158.19, 217.10
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.2546	1.2546	1.2546	1.3854	1.006
Resolution (Å)	20.4(4.5-4) ^b	20.4-2(5- 4.2)	20.4-5(5- 4.5)	20.4-5(5- 4.5)	20.4-5(5- 4.5)
<i>R</i> _{meas} (%) ^c	9.6(192.3)	11.3(137.5)	9.2(133.6)	18.8(194.4)	13.4(226.2)
<i>I</i> / σ <i>I</i>	11.2(1.0)	8.92(1.32)	9.04(1.34)	8.03(1.91)	8.78(1.02)
CC _{1/2} ^d	99.9(83.7)	99.9(91.1)	99.9(90.6)	99.8(95.9)	99.9(81.3)
Completeness (%)	98.2(97.8)	98.3(98.7)	98.4(99.3)	97.0(96.2)	98.0(98.7)
No. reflections	24400	21583	17703	17858	17321
Redundancy	6.8(6.6)	6.9(7)	6.9(7.1)	13.1(13.2)	6.9(7.1)
Anisotropy ^e	4.0 x 5.2 x	4.2 x 5.7 x	4.5 x 6.1 x	4.5 x 6.1 x	4.5 x 6.1 x
(<i>a</i> , <i>b</i> , <i>c</i>)	4.0 Å	4.2 Å	4.6 Å	4.5 Å	4.6 Å

^aDerivative 5, 1 crystal; derivative 6, 1 crystal; derivative 7, 1 crystal; derivative 8, 2 crystals; derivative 9, 1 crystal.^bHighest resolution shell is shown in parentheses.^cRedundancy of independent *R* factor calculated in XDS.^dPercentage of correlation between intensities from random half-datasets, as calculated in XDS.^eCalculated anisotropy from UCLA diffraction anisotropy server.

Extended Data Table 3 | Data collection and refinement statistics

	Native 2 ^a
Data collection	
Space group	C222 ₁
Cell dimensions	
a, b, c (Å)	88.18, 154.81, 219.77
α, β, γ (°)	90, 90, 90
Wavelength (Å)	1.000
Resolution (Å)	38.7-2.7(3-2.7) ^b
R_{meas} (%)	15.9(928.4) ^c
$I/\sigma I$	10.0(0.23)
CC _{1/2}	99.9(54.4) ^d
Completeness (%)	98.7(98.1)
No. reflections	41161
Redundancy	7.4(7.5)
Refinement	
Resolution (Å)	38.7-2.87 (2.8 x 4.0 x 3.3) ^e
No. reflections	21275
$R_{\text{work}}/R_{\text{free}}$	0.2970/0.3394 ^f
No. atoms	5402
Protein	5282
Ligand/ion	63
Water	57
B-factors	
Protein	110.50
Ligand/ion	123.2
Water	75.80
R.m.s deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.744

^aNative 2 data collected from a single crystal.^bHighest resolution shell is shown in parentheses.^cRedundancy of independent R factor calculated in XDS.^dPercentage of correlation between intensities from random half-datasets, as calculated in XDS.^eElliptically truncated data used for final refinement.^f5% of reflections were omitted from refinement for the calculation of R_{free} .

CAREERS

DIVERSITY A chemist helps others to come out and network **p.265**

INTERVIEWS How to answer the most-dreaded questions go.nature.com/zrv65y

 **NATUREJOBS** For the latest career listings and advice www.naturejobs.com

TARIKVISION



GROUP DYNAMICS

A lab of their own

The make-up of a lab is crucial to success in publishing its research — and now, scientists are exploring how to compose the best research group possible.

BY CHRIS WOOLSTON

Scientists around the world are working to solve the same basic formula: what number and mix of group members makes for the most efficient and productive lab?

Some principal investigators (PIs) produce a steady stream of high-impact papers with just a couple of people in the lab; others successfully oversee a team that could populate a village. Some stock up on postdocs, and others aim to balance career stages and positions: graduate students, staff scientists and technicians.

One of the most important steps for new PIs to take early in their career is to identify the formula that works best for them. In the past, they have had to decide the make-up of their group largely on the basis of their instincts and, often, financial realities. But now, they have some data to turn to. Studies on how lab size and

composition affect productivity give researchers guidance in their quest for better science, more publications and higher impact.

Junior faculty members who are deciding how to staff their lab need to consider their priorities: do they want to maximize the number of publications, or focus instead on impact? Do they favour hands-on or hands-off management? The number and type of people in a lab can affect all these important parameters, so PIs should build their labs with care — and with a plan.

BIGGER IS BETTER

Two studies published last year suggest that most labs could produce more papers and make a bigger splash by — perhaps unsurprisingly — bringing more people on board. One of these, a 2015 study of nearly 400 life-sciences PIs in the United Kingdom, found that the productivity of

a lab — measured by the number of publications — increased steadily, albeit modestly, with lab size (I. Cook, S. Grange, & A. Eyre-Walker *PeerJ* <http://doi.org/bcwf>; 2015). In terms of sheer paper production, “it’s best for a lab to be as big as possible”, says co-author Adam Eyre-Walker, a geneticist at the University of Sussex, UK. Notably, the study found no sign that individual members become less productive or less efficient as labs grow. “Adding a team member to a large lab gives you the same return as adding one to a small lab,” Eyre-Walker says.

The second paper, a study of 119 biology laboratories from 1966 to 2000 at the Massachusetts Institute of Technology in Cambridge, found that productivity inched forward when an average-sized lab of ten members added people (A. Conti & C. C. Liu *Res. Pol.* **44**, 1633–1644; 2015). But this study did detect limits: once lab size reached 25 people — an ▶

► unusually high number achieved by very few labs — the addition of team members no longer conferred benefit. Further, a lab's productivity tops out with 13 postdocs, the study found.

Co-author Christopher Liu, a former biochemist who now researches strategic management at the University of Toronto, Canada, points out that his study was limited to biology labs at one institution, which makes it tricky to generalize the findings. Still, he says, PIs should pay attention to the take-home message: bigger isn't always better. "Going from 15 to 20 people is probably not great," Liu says. "But going from two people to seven is something that you should probably do. A group of two people is pretty fragile."

GROWING PAINS

Sarah Teichmann, a molecular biologist at the EMBL-European Bioinformatics Institute and at the Wellcome Trust Sanger Institute in Hinxton, UK, can attest to both the pay-offs and the challenges of growing a lab. "When I started in 2001, it was just myself and a PhD student," she says. "I grew my group slowly. After three years, I had three PhD students and a postdoc."

She might have kept that modest configuration, but a change of focus forced a change in lab size. After several years of work on the computational aspects of gene expression and protein folding, Teichmann added an experimental angle to her research. She started by hiring one postdoc to focus on experimental work, but soon realized that he needed help. "He was alone and isolated," she says. "It didn't really work. There has to be a critical mass of experimental and computational people or it won't take off."

A €1.3-million (US\$1.4-million) grant from the European Research Council in 2010 enabled her to add three people, and her lab was on the way to bigger things, including more grants, awards and high-impact publications. Today, she leads a group of five postdocs, four PhD students and two staff scientists — one for the computational side and one for the experimental side — with a steady flow of visiting scientists.



Adam Eyre-Walker prefers in-depth discussion.

MATTERS OF SIZE

How to pick the right group

Lab size affects not only the principal investigator (PI), but also the other members of a research group. Postdocs and graduate students should think about the scope and scale of a lab when choosing a place to work, says Koen Venken, a geneticist at Baylor College of Medicine in Houston, Texas.

Venken says that large labs have much to offer trainees, including plenty of opportunity for independence. The PIs in such labs simply won't have time to look over everyone's shoulder. But that does not mean that team members will be left to their own devices. They have each other, and they can often call on lab technicians for help with tricky tasks.

Small labs might be better for trainees

who want a close, collaborative connection with their PI, Venken says. And, he adds, postdocs and graduate students who don't have the luxury of handing tasks over to a lab tech may end up learning skills that could be valuable in future job searches. Papers from small labs can be as important and influential as those from large labs. Furthermore, Venken notes, papers from small labs are less likely to carry a large roster of authors, which makes it easier for an individual contributor to stand out.

Ultimately, it is up to lab members to make the most of their situation, no matter where they land. "If someone is very proactive and innovative, they can be highly independent in a small lab, even when the PI is hands-on," Venken says. **C.W.**

Staying on top of such an enterprise has been daunting for her and her team (see 'How to pick the right group'). "The bigger your group is, the less face-to-face time you're going to have," she says. "There are only 24 hours in a day." Teichmann tries to keep the lab running smoothly by hiring people who work well together and support each other without her constant involvement. Her strategy is working: she has had her name on 16 publications since the start of 2015, including two articles in *Science*. She also won the 2015 EMBO Gold Medal, a prize awarded to outstanding young scientists in the life sciences.

Still, as Liu points out, bigger labs aren't always the key to a productive career. A smaller group can work for those who prefer to manage team members themselves and whose research doesn't require a huge roster.

For his part, Eyre-Walker finds comfort in the knowledge that small labs can make a big splash: his study found only a weak correlation between lab size and the average impact factor of each paper. He oversees a relatively small team of three PhD students and a postdoc, and says that he can remain deeply engaged with the analysis of all the work in his lab. "I couldn't cope with any more people," he says. "I like it like this. I can still do science. I'm not just managing people."

Some PIs learn through experience that they prefer a less-populous team. Koen Venken, a geneticist at Baylor College of Medicine in Houston, Texas, rapidly built a team of ten lab members after starting his faculty job in 2014. But he soon realized that his team members weren't working well together. "It was a mistake, and I'm happy to admit it," he says. After some rapid downsizing, he now has a group of two PhD students, one postdoc, one lab technician, one research associate and a non-tenure-track instructor, a mix that

has proved to be productive and efficient.

Looking back, Teichmann is glad that she took a slow, deliberate approach to building her lab. "Going slow is important for the sanity of the PI," she says. Eyre-Walker agrees. "You have to feel your way into it. Start small, and see how you get on. The worst thing you can do as a new faculty member is take on five PhD students."

QUALITY FIRST

But size is only one measure of a lab. PIs who are assembling a team also have to consider balance — and that means weighing the relative merits of graduate students, postdocs, technicians and other potential members. According to the *Research Policy* study, postdocs — especially those who have external funding through fellowships — are the key drivers of productivity. Overall, adding a funded postdoc to the average lab boosts output by about 29% of a published paper every year.

Graduate students don't contribute much to productivity, but they do play an important part in the group. The analysis found that students are as valuable as funded postdocs for generating 'breakthrough' papers, which the study defined as anything published in *Science*, *Nature* or *Cell*. Adding either a funded postdoc or a graduate student to the average lab increases the chances of such a paper by about 8%, the team found. Postdocs without their own funding, who may not be quite as accomplished as their funded peers, do not improve the odds of a breakthrough paper at all.

Many PIs eventually have to concede to financial and other realities. Sergey Kryazhimskiy, an evolutionary biologist at the University of California, San Diego, was originally dead set against hiring postdocs. He recognized that many postdocs end up stuck in their positions and are not able to move on to tenure-track

jobs — and he did not want to play a part in what he views as an unfair system with enormous stakes. “If you’re a responsible PI, you would like your postdocs to proceed somewhere after your lab,” he says. “It’s difficult to assign them risky projects. You’re playing with their lives.”

He had a plan for avoiding his ethical dilemma: he would bring in staff scientists who were committed to their lab careers. But when he actually got his faculty position earlier this year, he realized that pragmatic considerations outweighed the ethical ones. He estimates that at his institution, it costs nearly twice as much to hire staff scientists as it does to hire postdocs, partly because they get benefits such as paid time off and health insurance.

Unable to stick with his original strategy, Kryazhimskiy has started to interview postdocs. He is looking for candidates whom he thinks will have a good shot at a faculty job, even in a tough academic market. Another option is to find someone with other career goals, such as a job in industry. From a purely practical perspective, he thinks that postdocs will be the best investment of his grant money.

PIs whose labs — and grants — are on the large side may be better able to absorb the cost of staff scientists. For Teichmann, at least, her two staff members are key to her lab’s success. Both are accomplished researchers who know how the lab works and how to get things done. She expects to hire two more professionals: a lab manager and a software developer. “Then I would have four core people who can support my postdocs and PhD students,” she says. Unlike postdocs and graduate students, those four professionals wouldn’t be locked into a pressurized timeline to graduate or to move on to another job.

Venken would eventually like to add a few people to his lab, too — perhaps some postdocs, graduate students or a mixture of both. “I just want people who are invested in everything that we’re doing,” he says.

The size and structure of a lab can be hugely important, but in the end, the quality of any workplace comes down to the quality of the people, PIs say. Whether they are looking for graduate students or postdocs, whether they desire a large or small research group, new PIs need to find team members who are ready to contribute. “The first set of individuals that you hire is very important,” Liu says. “They set the tone for the entire laboratory.” ■

Chris Woolston is a freelance writer in Billings, Montana.

TURNING POINT

Out for chemistry

David Smith, a chemist at the University of York, UK, spent his early career avoiding personal discussions with colleagues because he did not want to reveal that he is gay. In January, he gave the plenary talk at the first LGBT (lesbian, gay, bisexual, transgender) STEMinar, a conference devoted to networking.

How did the LGBT STEMinar come about?

A postdoc at the University of Sheffield, UK, Beth Hellen, decided that she wanted to get a bunch of LGBT scientists she knew through Twitter together for networking. She thought 20 people would attend, but about 80 showed up. It was, as far as I know, the first ever meeting in the United Kingdom to specifically target LGBT scientists across all disciplines. It was a really nice meeting, with genuine networking. Similar things have gone on in the United States, especially at the big conferences, like the American Chemical Society meetings. But this has never been a feature of UK–European science.

Do you think it will continue?

Yes. One of the most heartening things about the meeting was that it got support from high-level societies such as the Royal Society of Chemistry and the Institute of Physics. It’s a time of big change in science. Fifteen years after the culture broadly changed, we are now talking about our personal lives and acknowledging who we are. There are plans for another LGBT STEMinar at Sheffield next year.

How did you find the diversity as a student?

It was not great. I think when I was at the University of Oxford, UK, where I got my PhD, there were about 1,000 chemists in total. At least 75% of them were white men. I have no idea how many of the chemists were LGBT, but I do know that they were silent. Occasionally, there were rumours or gossip about individuals, but it was always negative. It was a hostile environment in the early 1990s. That started to change when former prime minister Tony Blair introduced civil partnerships in 2004.

So ‘don’t ask, don’t tell’ was the de facto policy?

Yes. I wasn’t ‘out’ when I started at the University of York. As a result, I engaged in a lot of self-censorship. When chatting about the weekend with colleagues, I’d neutralize the gender of my partner or just not talk about my personal life at all. But I’d end up in difficult situations — half lying, half telling the truth and trying to remember what I had told individual people to be consistent in conversations.



JOHN HOULIHAN

What prompted you to come out?

I was in a long-term relationship and it got more ridiculous not to talk about it. I had been in my job for 4 or 5 years when another gay colleague arrived in the department. It gave me a bit of confidence. I came out in 2002, and I received an overall positive response. Some people were surprised but the uncomfortable period didn’t last long. York has one of the most diversity-friendly chemistry departments.

You’ve been very open since then. Do junior colleagues contact you to discuss LGBT issues?

Yes, I get tens of e-mails from people globally, often people in junior positions, such as postdocs who are unsure about what impact coming out could have on their career. The apprenticeship model leaves junior researchers dependent on their supervisor’s recommendation. People worry that even unconscious bias could bleed into a reference letter for a job application. There’s no easy answer. Every supervisor is different. The last thing I want to do is say ‘come out’, and have supervisors write horrible letters.

You make fun YouTube videos, and encourage your students to do so, too. Why?

My videos — notably the chemistry of mephedrone or the science behind the television show *Breaking Bad* — got general traction beyond students. I decided to encourage my students to make videos as a way to empower them with a voice. I wanted them to realize that they don’t have to just absorb knowledge, they can be a source of it. It also became a way for me to discuss diversity issues and use it as an education tool. ■

INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.

SHOVELWARE

A fresh connection.

BY BOGI TAKÁCS

Tamás first bumped into his next-door neighbour while carrying a box full of kitchen utensils upstairs to his second-floor apartment. He cursed himself for not renting a utility drone to fly his boxes in through the window.

"Hey," his neighbour said. "Can I help?"

He couldn't see her — the box was blocking his view. "Thanks, I'll manage." Pots and pans clanged together as he slipped on a stair worn concave from use. She chuckled, then grabbed the box.

He only got a good look at her once they put down the weight. She was tall, muscular, ethnically mixed. She was wearing college sports fatigues, a faded black T-shirt with a tech company logo and a paisley pattern headscarf.

"Salaam," Tamás offered.

"Wa alaykum," she said, "but I'm not Muslim. Just wearing this because my bald head gets cold."

Cancer? He didn't dare ask. "Nice to meet you. I'm Tamás."

"Liliane."

After five more boxes, she invited him over for tea.

"You deserve this," she said, spooning honey into his cup. "You're so thin!"

"Intercontinental move," he sighed.

"Where are you from?"

"Hungary."

"Oh." She fell silent for a moment. "I saw it on CNN. I'm sorry. Here, have one of my sandwiches. You have family back there?"

Fresh tomatoes and salad crunched under his teeth. He swallowed hard. "No family."

"I'm glad you got out in time. What's your line of work?"

"I'm a painter. Oils, the occasional watercolour, some digital stuff. Yours?"

"I dream video games." She smiled.

"Design lead? Concept artist?"

"No — I dream them. After you're done, I can show you my rig."

He wiped his hands on his paint-stained trousers.

"Just to be sure," she said, "letting you into my bedroom has no sexual connotations. Don't get too excited."

He shrugged. "I'm gay."

She opened the door. "Good. I'm just happy to talk shop with a fellow artist, you know?"

One wall had a French window — the other held a tangle of equipment.

"Is that like one of those imaging things, Mind's Eye?" He'd contemplated buying one, but he was wary — he knew the devices were non-invasive, they just recorded and displayed mental imagery based on signals registered on the scalp, but he felt there was a limit to how close he wanted to be with technology. He wouldn't shave his head.

"Yeah, with a bit more resolution than the consumer models," she replied. "I'm testing this one for the company now." He connected the dots — the logo on her T-shirt.

He nodded. "But I don't get it. Why do you do it asleep?"

"You know lucid dreaming? I can control my dreams. Natural talent, I guess. Discovered I could do it as a kid. But some people can also learn it." She grinned at him.

"Why do you need to dream for that?"

"Much faster. I can dream entire games per sleep cycle. Then the coding team just needs to export the art, the music, code the rules, et cetera. All the art assets in one sitting. I can't hold all that in my head while I'm awake, but my brain takes care of it while I'm asleep. I'm best at jump and run, platforming, those types of games."

He grinned back. "Cool. Do you develop for consoles? PC?"

"We just push them out to mobile app stores," she grimaced. "Shovelware, you know the term." He didn't, but he could understand. "We make 'em by the truckload."

"I used to do serigraph prints. Not the same, but I get the point." Not the same at all. He felt disillusioned — he'd expected something glamorous. Dreaming video games! Then again, it seemed that the goal was to speed up the development process, not to improve on it.

"Not so exciting when I put it this way, huh? I'd love to make a survival horror adventure sometime. They say it's not my genre."

Horror he knew about. "Giger was one of my major influences," he said. "That and politics. I can show you my art. Tomorrow after unpacking?"

He felt like he was making her depressed.

Was she making him depressed? He bid her goodbye, then spent the rest of the day scrubbing his kitchen.

He showed her his art. They made more tea and moped in her comfy couch.

Then they avoided each other for weeks.

He found out she was freelancing for various shovelware companies. He got himself a new phone, busied himself with apps. There was a new one every day, her energetic demeanour all over them.

Then they started to turn gloomy.

It took him a month to realize she must've been looking at his art. And maybe more. Was that building in the background the Hungarian Parliament, burning? It scrolled by so fast.

He knocked, a tray of cookies in hand. "I'm sorry. I just thought you might..."

"Come on in!" She was cheerful.

They munched, a shared love of cooking creating a bond between them. "I was looking at your games."

"I was looking at your art," she said, unfazed.

"I know," both of them said at the same time. They laughed.

"Fancy a collaboration?" she asked, then held up her hands. "Not kidding! A friend's making a leap, setting up a start-up. We could make that dark game I've been dreaming about."

He'd also been dreaming about it. She held no romantic attraction to him, but they intuitively meshed as friends. As collaborators?

She leaned forward. "Tell me about Hungarian politics."

He ranted on and on, years of frustration finally allowed an outlet. He was safe. He could create — without self-censorship, without doubletalk, without shame.

Outside, the Sun slowly set. ■

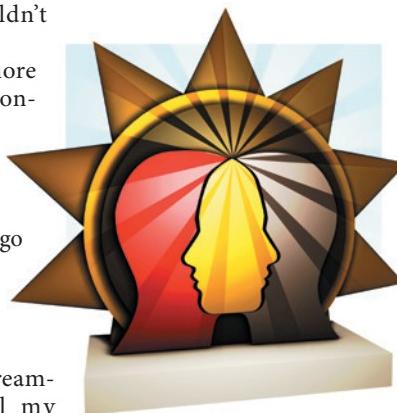


ILLUSTRATION BY JACEY

Bogi Takács works in the University of Iowa Word Learning Lab by day and writes stories by night. Bogi's speculative fiction and poetry have been published in venues such as Clarkesworld, Lightspeed, Strange Horizons and Apex.